



**ПРИМЕНЕНИЕ МНОГОМЕРНОГО
СТАТИСТИЧЕСКОГО АНАЛИЗА
В ЭКОНОМИКЕ И ОЦЕНКЕ
КАЧЕСТВА ПРОДУКЦИИ**



ТЕЗИСЫ II

АКАДЕМИЯ НАУК СССР
ЦЕНТРАЛЬНЫЙ ЭКОНОМИКО-МАТЕМАТИЧЕСКИЙ
ИНСТИТУТ
МОСКОВСКИЙ ЭКОНОМИКО-СТАТИСТИЧЕСКИЙ
ИНСТИТУТ
ТАРТУСКИЙ УНИВЕРСИТЕТ

**IV ВСЕСОЮЗНАЯ
НАУЧНО-ТЕХНИЧЕСКАЯ КОНФЕРЕНЦИЯ**

**« ПРИМЕНЕНИЕ МНОГОМЕРНОГО
СТАТИСТИЧЕСКОГО АНАЛИЗА
В ЭКОНОМИКЕ И ОЦЕНКЕ
КАЧЕСТВА ПРОДУКЦИИ »**

Тезисы докладов

II часть

5—7 сентября 1989 года
г. Тарту (Эстонская ССР)

ТАРТУ 1989

Ответственный редактор Э.-М. Тийт

Anh.

Tartu Ülikooli
Raamatukogu

10444

KUSTUTATUD

IV ВСЕСОЮЗНАЯ НАУЧНО-ТЕХНИЧЕСКАЯ КОНФЕРЕНЦИЯ
"ПРИМЕНЕНИЕ МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА
В ЭКОНОМИКЕ И ОЦЕНКЕ КАЧЕСТВА ПРОДУКЦИИ".

Тезисы докладов. II часть.

На русском языке.

Тартуский университет.

ЭССР, 202400, г.Тарту, ул.Ülikooli, 18.

Ответственный редактор Э.-М. Тийт.

Подписано к печати 28.07.89.

МВ 03106.

Формат 60x84/16.

Бумага ротаторная.

Машинопись. Ротапринт.

Условно-издательских листов 9,67.

Условно-печатных листов 10,00.

Печатных листов 10,75.

Тираж 400.

Заказ № 569.

Цена 65 коп.

Типографию ТУ, ЭССР, 202400, г.Тарту, ул.Тийги, 78.

СОДЕРЖАНИЕ

<u>Программное обеспечение</u>	<u>Экспертные системы</u>
С.В.Белоусова, А.А.Дыховичный, С.В.Ефименко, И.И.Клесов, О.И.Клесов, В.И.Колчинский, Т.В.Кравец, Р.Е.Майборода, О.С.Пилявская, Д.С.Сильвестров, О.Ю.Хайдуков	
Автоматизированная система первичного анализа АСПА	188
В.П.Бурдаев	
Интеллектуальная система кластер анализа дан- ных	190
Е.Е.Витяев, Н.А.Подкольный	
Научно-исследовательская экспертная система	192
А.Гавсюк, К.Губин, Г.Гамбаров, А.Майданчик	
Экспертная система классификации "ЭСКИМО"	194
Е.В.Маркова, Н.Г.Чеботарев	
Основные и вспомогательные компоненты методо- ориентированной системы ПШАНЭС	195
М.Д.Степанова, Т.И.Колесник	
Один из подходов к созданию экспертной систе- мы для статистического консультирования	197
<u>Программное обеспечение</u>	<u>Универсальные системы</u>
С.Ю.Адамов, М.К.Вегнер, А.А.Халилеев, А.Г.Шимельмиц	
Система анализа нечисловой информации "С А Н И" ..	199
Х.Буш, Р.Полис, О.Штродах, А.Клейн	
Комплексирование программных средств на базе системы статистического анализа	201
С.А.Егоров	
Системная интеграция прикладных программ на базе расширенной концепции гипертекста	202

И.С.Енюков, Н.Д.Вылегжанин, А.С.Соломоник	
Система интерактивного анализа многомерных данных (СИАМД)	204
Ю.М.Кабанов, М.А.Седова, Г.В.Срагович	
Статистическая система "Эконометрика"	206
А.В.Калита	
Об опыте разработки программного обеспечения анализа временных рядов	207
М.Каролин, Т.Тийтс	
СТАТЕС-4 со стороны пользователя	209
М.Каролин, Т.Тийтс, В.Ольман	
О некоторых дополнительных возможностях системы СТАТЕС-4 по сравнению с пакетом ВМОР при анализе нелинейной регрессии и эмпирических распределений	211
Т.Кинкар	
Модифицированный "ПАКСТАТ"	213
В.О.Мазур, А.Ю.Хровский, А.М.Никифоров, Е.Г.Ардаева, А.С.Титов, К.М.Богатов, Е.Е.Кожаринова, С.В.Углова, С.В.Сливко-Кольчик	
Пакет статистических программ "Анализ данных" ...	214
И.Н.Парасюк, О.В.Вережка, А.В.Калита, Е.С.Карпинка, А.И.Провотарь, Н.П.Скопецкая	
Система СИГМАСТАТ: функциональные и инструментально-технологические возможности	216
Т.Тийтс, М.Каролин	
Интегрированная система обработки статистических данных - СТАТЕС-4	218
<u>Программное обеспечение</u>	<u>Метод-ориентированные системы</u>
Ю.Вилисмяэ	
Интерактивный режим классификации в СВМ	220
В.И.Елейко	
Об одном подходе к прогнозированию рядов динамики	221
Е.И.Жилинская	
О библиотеке программ построения доверительных интервалов при оценивании величины эффек-	

та воздействия	222
В.Йоала	
Быстрый метод оценивания полиномиального трен- да временного ряда	223
С.Е.Кузнецов, Д.В.Левандо	
Экспертная поддержка выбора модели Бокса-Джен- кинса в системе анализа временных рядов МЕЗОЗАВР ..	224
С.Е.Кузнецов, А.А.Халилеев	
Статистическая система по анализу временных рядов "МЕЗОЗАВР"	226
Т.М.Пестунова	
Пакет программ "ЛОГР" для решения задачи груп- пировки объектов	228
Т.А.Пригарина	
Программное обеспечение анализа парных срав- нений	230
Г.А.Сатаров, Е.В.Кулинская, Г.В.Раушенбах	
Специализированный пакет C L A M S : класси- фикация и многомерное шкалирование	232
В.И.Сердобольский, В.С.Степанов	
О комплексе программ ДИАНА - Дискриминант- ного Анализа Векторов Высокой Размерности	234
Н.Н.Товмаченко, О.В.Майстренко, О.С.Малюга	
Программное обеспечение методов главных ком- понент и дискриминантного анализа в задачах большой размерности	236
М.Хоолма, Э.Лелумезс, Э.Вийкман, И.Петерсен	
Пакеты КЛАССИСТАТ и КЛАССИСТАТ-Д для статис- тического анализа данных	237
К.Чепонис	
Экспериментальное исследование работы класси- фикаторов при применении нормализующих транс- формации	238
<u>Программное обеспечение</u>	<u>Проблемно-ориентированные системы</u>
В.О.Гречко, И.В.Жовтяк, А.М.Захарин	
Проблемно-ориентированные процедуры статисти- ческого анализа показателей средств информатики ..	240

В.А.Лапидус	
Байесовский подход к статистическому контролю качества при использовании принципа распределения приоритетов	242
А.П.Макаров, Ю.К.Беллев, В.А.Беллева, Т.Н.Дугина	
Комплекс прикладных программ по анализу цензурированных данных (POD)	244
Е.Н.Наумова, Т.В.Елисеева, В.В.Губарев, М.П.Цапенко, И.Б.Белан	
Применение статистического анализа в клинической иммунологической лаборатории	246
В.Пиворюнас	
Интеллектуальная система ARFA (Автоматизированный Рентгенофазовый Анализ)	248
Г.В.Раушенбах, В.Д.Воскобойкин, А.А.Заславский, В.Б.Митрофанов	
Математическое обеспечение многомерного статистического анализа экспертной и экономической информации при формировании государственного заказа	250
<u>Оценивание математического обеспечения</u>	<u>Тестирование</u>
Ю.А.Великий	
Обзор популяции отечественного статистического программного обеспечения	252
М.Вяхи	
Тестирование процедур дисперсионного анализа методом точных выборок	253
В.П.Морозов, И.П.Колпаков	
Система экспертно статистической оценки качества программных средств (ЭСОК ПС)	255
М.Л.Петрович, М.И.Давидович, Т.И.Левина, Д.А.Мацкевич, А.Н.Шитик, Г.К.Шлег	
Численное исследование алгоритмов прикладной статистики	256
В.Пикалис	
Исследование статистических характеристик экспериментального выбора оптимального алгоритма обработки данных	257

Д.С.Сильвестров	
Методика экспертного обследования статистического программного обеспечения	258
Э.Тийт	
Тестирование матобеспечения по статистике - теоретические и методические проблемы	259
<u>Оценивание математического обеспечения</u>	<u>Экономика обработки</u>
А.Н.Гладун, Б.М.Кац	
Многомерный статистический анализ экономики отрасли информационно-вычислительного обслуживания	260
Л.М.Тоодинг	
Эмпирические параметры распределения в реальном потоке статистических данных	262
Н.А.Цибель	
Соотношение качества и цены прикладного программного обеспечения	264
<u>Приложения</u>	<u>Экономика</u>
В.А.Адамчик	
Применение методов многомерного анализа в изучении эффективности машиностроения СССР	266
Е.Н.Батугина	
Многомерный статистический анализ в угольной промышленности	268
И.Б.Береснева	
Исследование закономерностей развития региональных систем областного ранга классификационно-типологическими методами	270
В.В.Богачев, И.П.Колпаков	
Прогнозирование временных рядов экономических показателей с использованием ППП ОЛИМП и СКИФ	272
И.М.Денисова, О.А.Андрюшкевич	
Разработка системы моделей мирохозяйственных связей с реализацией адаптивного подхода	273
Н.И.Киселев	
Оценивание тренда параметров в эконометри-	

ческих зависимостях	275
А.Б.Колбасова	
Прогнозирование спроса на особо модные изделия в легкой промышленности	277
И.Н.Минскер	
Статистическое исследование адекватности оценки научно-технического уровня АСУТП	279
В.Ольман	
Задача о выборе типопредставителей в обувной промышленности	281
И.В.Семянк, Е.И.Березань	
Многомерный статистический анализ покупательской мотивации выбора товара	283
И.В.Семянк, П.А.Иващенко	
Применение методов кластеризации и ранговой корреляции в изучении тенденций развития промышленных предприятий	285
<u>Приложения</u>	<u>Общественные науки</u>
С.Ю.Адамов, И.А.Герасимова	
Моделирование лонгитюдных данных	286
В.А.Балаш, Н.Ф.Горчакова	
Моделирование взаимосвязи показателей социальной активности работников АПК	288
Н.В.Высоцкая, О.В.Терещенко	
О применении многомерного статистического анализа в изучении демографической ситуации	289
И.Р.Гиль, Я.П.Лумельский	
Планы контроля по качественному признаку и характеристики регламентов матчей	291
Т.Р.Минина, Т.В.Маслякова, Е.В.Фример	
Прогнозирование социально-экономических процессов в рамках классификационно-типологических моделей	293
А.А.Митоян	
Методы целенаправленного проецирования в выявлении типологии потребительского поведения семей	295

И.Ф.Петерсен

Применение многомерного статистического анализа к прогнозированию развития и к исследованию стабильности многопартийных политических систем ... 297

О.В.Староверов

Статистические потенциалы 299

Т.В.Хачатурова, Е.М.Ильин, В.Т.Перекрест

Средства типологического моделирования социально-экономической информации 300

Приложения

Экология и медицина

В.Н.Виноградова, А.Г.Закиров, Л.Л.Фролова

Автоматизированное рабочее место эколога-исследователя 302

Е.А.Зеленюк

Статистическое исследование компонент загрязнения атмосферы 304

А.А.Зубенко, Г.М.Зубенко

Применение универсального по ПС для проблемно-ориентированного статистического анализа данных о состоянии окружающей среды 306

А.Б.Качинский, В.И.Польченко, Н.П.Сарническая

Множественный линейный регрессионный анализ влияния химических классов пестицидов на здоровье населения 308

А.Маамяги

Анализ экспертных мнений о причинах и мерах предохранения Чудского озера от загрязнения 310

А.Маамяги, С.Малыгин

Статистическая обработка экспериментальных данных фосфоритных месторождений Тоолсе и Кабала 311

Т.Мелс

Оценивание глобальных статистических параметров молекулярной эволюции 312

О.И.Рябкова, И.Б.Марченкова

Принципы организации статистической обработки медико-генетических данных средствами прог-

рамной системы ОМЕГА	313
П.Н.Сапожников	
Планы блужданий полиномиального типа в моде- лях нарушения ритма сердца	315
С.Т.Трусенков, О.О.Трусенкова	
Автоматизированная методика выделения естест- венных гидрологических аномалий в целях эко- логического мониторинга	317
Г.Л.Шевляков	
Статистическая классификация уровней внезап- ной смертности в зависимости от метеофакторов	319
<u>Приложения</u>	<u>Проверка качества, стандарты</u>
Н.В.Алтухова	
Комплексный многомерный анализ эффективности обслуживания оборудования	320
А.П.Демиденко	
Нечеткие множества и нечеткая статистика в за- дачах оценки качества	322
В.Г.Кобзев, Н.Г.Николенко, А.А.Репко	
Применение методов анализа стабильности значе- ний информативных параметров для оценки качес- тва электрорадиоизделий	323
Г.Г.Кравченко, А.И.Орлов	
Метод выделения единиц бесформенной продукции при статистической контроле качества	324
Т.К.Курочкина	
Применение многомерного статистического анали- за для ранжирования причин нарушений качества работы комбайнов	325
А.В.Медведев, И.А.Пестунов	
Непараметрический алгоритм оценки качества продукции при изменяющихся условиях	326
В.С.Мхитарян, С.В.Чуприн	
Многомерный статистический анализ качества и надежности прокатных валков	328
А.И.Орлов, О.М.Черномордик	
Проект "СТАТПРОМ"	329

И.Б.Руссман, В.М.Умывакин	
Анализ данных и интегральные оценки качества при проектировании сложных систем	331
А.П.Рыжаков, А.Ф.Терпугов	
Об оценке влияния стабилизации факторов на ка- чество продукции	332
С.В.Соболев	
Новое в оценке качества текстильных материалов ..	334
О.И.Тескин, С.Я.Криволапов	
Интегральное оценивание показателей надежности по результатам форсированных испытаний	336
Ю.Я.Чижиков, Н.Б.Якимова	
Применение методов многомерного шкалирования в задачах оценки качества операционной регули- ровки реле	338
Н.Н.Швайдак, Е.Н.Румянцев, С.В.Гурин	
Терминальные модели в задачах управления ка- чеством продукции	339
Авторский указатель	341

АВТОМАТИЗИРОВАННАЯ СИСТЕМА ПЕРВИЧНОГО АНАЛИЗА АСПА

С.В.Белоусова, А.А.Дыховичный, С.В.Ефименко, И.И.Клесов,
О.И.Клесов, В.И.Колчинский, Т.В.Кравец, Р.Е.Майборода,
О.С.Пилявская, Д.С.Сильвестров, О.Ю.Хайдуков

Автоматизированная система АСПА предназначена для статистической обработки с помощью ЕС ЭВМ информации, представленной временными рядами большой длины, с целью отыскания разладок, т.е. моментов изменения вероятностных свойств, а также аномальных участков временного ряда.

В рамках реализованной в системе АСПА автоматической технологии поиска разладок и аномалий исходный временной ряд преобразуется в многомерный временной ряд так называемых сжатых характеристик. Его членами являются векторы, содержащие вычисленные по секциям определенной длины различные вероятностные характеристики исходного ряда. К последним относятся: корреляционные и спектральные функции, функция распределения, различные выборочные статистики. дальнейшая статистическая обработка основана на анализе этого многомерного ряда сжатых характеристик и включает широкий спектр современных математических методов: последовательный и апостериорный поиск разладок, классификация, поиск информативных характеристик и т.д.

В системе АСПА реализованы различные методы визуального анализа многомерных наблюдений в интерактивном режиме. Пользователю предоставляется целый ряд сервисных возможностей, позволяющих просматривать интересующие участки ряда (в том числе в режиме мультипликации), фиксировать те или иные особенности данных, инициировать выполнение различных процедур полуавтоматического и автоматического анализа данных.

Характерной особенностью системы АСПА является возможность практически неограниченного расширения как функционального, так и системного наполнения. С целью поддержки такого пополнения системы АСПА пользователю предоставляются сле-

дующие возможности: динамический вызов алгоритмов; виртуализация переменных; динамическое управление оперативной памятью из программ на любом языке программирования; ведение базы данных; контроль и обработка аварийных ситуаций, обмен информацией с дисплеем ЕС ЭБМ.

В системе АСПА реализован специальный входной язык высокого уровня, по синтаксису сходный с языком ПАСКАЛЬ. Отличительной особенностью АСПА-языка является наличие специальной конструкции "настройка", которая позволяет пользователю описать конкретную конфигурацию системы, соответствующую целям и задачам пользователя.

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА КЛАСТЕР АНАЛИЗА ДАННЫХ (ИСКАД)

В.П.Бурдаев

Кластер-анализ данных - это выделение структур, классов, однородных групп (кластеров) из совокупности объектов, описываемых набором признаков. Он позволяет: получить разбиение объектов на некоторое число кластеров, снизить размерность пространства признаков и отобрать наиболее информативные признаки.

Системы, основанные на знаниях (экспертные системы), находят широкое применение при решении задач диагностики в самых различных предметных областях таких, как медицина, биология, геология, химия и другие.

Система ИСКАД предназначена для решения задач диагностики и автоматической классификации многомерных объектов с использованием знаний экспертов.

Она имеет многоуровневую иерархическую модульную структуру, снабжена инструментальными средствами для создания баз знаний и имеет дедуктивную машину вывода с объяснением, использующую обратную систему продукций для правил и прямую систему продукций для метаправил. Во время консультации генерируются кластеры базы знаний.

Сведения о применении алгоритмов автоматической классификации отражены в правилах и фактах, содержащихся в специальной базе знаний по кластерному анализу.

Система позволяет пользователю, не являющемуся специалистом в области вычислительной техники:

- сформировать и обслуживать локальную базу данных;
- редактировать числовые и символьные данные, тексты;
- выполнить предварительную обработку матрицы объект-признак;
- визуализировать данные;

- выполнить интерактивную кластеризацию;
- отобрать информативные признаки;
- выполнить интеллектуальную кластеризацию, используя соответствующую базу знаний;
- создавать и редактировать базы знаний: правила, факты, вопросы и ответы для атрибутов предметной области;
- выполнить консультацию, указав её цель;
- получить объяснения во время консультации;
- просмотреть маршрут работы машины вывода и получить протокол консультации;
- обработать экспертные оценки, провести экспертизу.

Дальнейшее развитие системы ИСКАД состоит в наращивании базы знаний по кластерному анализу, привлекая методы нечеткой кластеризации, и в разработки экспериментальной базы знаний в сельском хозяйстве.

Технические характеристики ИСКАД:

- система используется на персональных компьютерах с операционной системой MS DOS;
- система программирования - TURBO PASCAL 5.0;
- объем загрузочного модуля составляет 250 Кбайт;
- система снабжена инструкциями для работы с пользователем;
- система поставляется на двух флоппи-дисках: на первом размещается загрузочный модуль ISCAD, на втором - необходимые файлы содержащие: подсказки для работы с пользователем, контрольные примеры, базу знаний по кластерному анализу и демонстрационные базы знаний (например, диагностирование на ишемическую болезнь сердца);
- максимальная размерность матрицы объект-признак составляет 3000 элементов.

Замечание. С целью использования больших массивов данных или для проведения консультаций с большой базой знаний (свыше 400 правил) система может быть расчленена на отдельные модули, т.е. адаптирована к задаче пользователя.

НАУЧНО-ИССЛЕДОВАТЕЛЬСКАЯ ЭКСПЕРТНАЯ СИСТЕМА

Е.Е.Витяев, Н.А.Подколодный

Система ориентирована на специалиста-исследователя и предназначена для поддержки научных исследований. Система дает возможность специалисту-эксперту на основании имеющихся у него знаний; априорных знаний из литературных источников и данных, полученных в результате сбора информации (обследования, анкетирования, экспериментирования, тестирования и т.д.), построить "полную" в определенном смысле, непротиворечивую систему знаний - эмпирическую теорию предметной области [1].

Система является результатом развития концепции экспертных систем второго поколения в сторону экспертных систем-теорий путем добавления индуктивного синтеза правил, как в диалоге с экспертом, так и путем обнаружения "полного" множества логических закономерностей на имеющихся данных, а также определения "истинных" шкал величин и законов. Система может применяться в любой области деятельности, где есть данные, но нет теории, либо она слабо развита.

В состав системы входит:

- реляционная СУБД с языком SQL ;
- СУБЗ, поддерживающая продукционно-фреймовое представление знаний;
- интерфейс пользователя с подсистемой объяснений;
- блок формирования и тестирования базы знаний с подсистемой приобретения знаний;
- подсистема логического вывода и принятия решений;
- подсистема индуктивного вывода правил.

Главной особенностью системы являются развитые средства построения базы знаний. Система предоставляет возможности ввода и редактирования априорных знаний в виде произвольной совокупности высказываний в языке первого порядка. В диалоге с экспертом это множество высказываний приводится к совокупности правил вида "если - то".

Формирование системы правил выполняется методом индуктивного синтеза ПРОЛОГ-программ на основании имеющихся в БД

или предъявляемых экспертом истинных и ложных примеров, либо примеров правильного рассуждения.

Другой подход, использованный в системе, — формирование вероятностных правил методом индуктивного синтеза ПРОЛОГ-программ по выборкам из данных. Такие правила имеют статус закономерностей и находятся проверкой некоторых статистических гипотез [2].

Система выявляет и устраняет противоречия в теории на разных этапах ее построения. Проверка неполноты теории проводится в диалоге с экспертом путем выявления невозможности вывести некоторый указанный экспертом факт или утверждение.

Система позволяет осуществлять вывод из полученной теории систем аксиом репрезентативной теории измерений и аксиоматической теории принятия решений для определения "истинных" шкал величин и законов, а также вывод систем аксиом теории конструктивных моделей для получения конструктивных числовых представлений "структурных" величин и законов.

В результате работы системы конструируется экспертная система, позволяющая решать любые задачи, выводимые из построенной теории, в частности, задачи предсказания (диагностики, прогнозирования, заполнения пробелов и т.д.); интерпретации выводов, закономерностей, процедур шкалирования "истинных" величин и законов; планирования эксперимента и т.д.

Система разрабатывается на языке ПРОЛОГ для IBM PC/XT/AT (MS.DOS) и LAETAM (UNIX) .

Л и т е р а т у р а

1. Витяев Е.Е., Москвитин А.А. ЛАДА — программная система логического анализа данных // Методы анализа данных. — Новосибирск, 1985. — Вып. III: Вычислительные системы. — С.23-32.
2. Витяев Е.Е. Индуктивный синтез ПРОЛОГ-программ методом обнаружения закономерностей // Проблемы совершенствования синтеза тестирования, верификации и отладки программ: Тез. докл. — Рига, 1986. — С.65-66.

ЭКСПЕРТНАЯ СИСТЕМА КЛАССИФИКАЦИИ "ЭСКИМО"

А.Гавсюк, К.Губин, Г.Гамбаров, А.Майданчик

Построение теории, позволяющей связать классификацию с целью исследования, свойствами и требованиями к разбиениям дало возможность разработать экспертную систему классификации "ЭСКИМО". Она реализует вероятностную модель в рамках надежного подхода в теории классификации [1].

"ЭСКИМО" предназначена для разбиения совокупности объектов на однородные группы с учетом знаний специалистов в области анализа данных, экспертов, работавших с системой, и конкретного пользователя. Она является гибридной многоуровневой экспертной системой. Это связано с тем, что, во-первых, в теории классификации большую роль играют формальные методы; во-вторых, в целях повышения эффективности ее работы и расширения возможностей интеллектуального интерфейса, в "ЭСКИМО" используются разнообразные формы представления информации: продукционные правила, фреймы, текстовая, числовая, графическая.

"ЭСКИМО" решает задачи выяснения знаний пользователя об изучаемом явлении и его требований к результатам классификации; проведения классификации с учетом знаний и требований пользователя; интерпретации результатов; объяснения вывода; корректировки результатов на основе обсуждения их качества с пользователем и внутренних формальных критериев; накопления знаний для повышения эффективности решения последующих задач.

"ЭСКИМО" состоит из базы знаний и машины вывода. База знаний включает знания трех типов: фактические, процедурные и сервисные. Машина вывода состоит из блока управления, интерпретатора правил, системы управления базой знаний, интеллектуального интерфейса, блока формального вывода и блока логического вывода.

Первая версия "ЭСКИМО" ориентирована на экономическую предметную область и апробирована на ряде практических задач.

Л и т е р а т у р а

1. Гамбаров Г.М., Майданчик Ар.Б. Теория классификации. - М.: деп. в ВИНТИ. - 1988, 350 с.

ОСНОВНЫЕ И ВСПОМОГАТЕЛЬНЫЕ КОМПОНЕНТЫ МЕТОДО-ОРИЕНТИРОВАННОЙ СИСТЕМЫ ПЛАНЭКС

Е.В.Маркова, Н.Г.Чеботарев

Интегральная система ПЛАНЭКС предназначена для консультативной и справочной службы в области математических методов планирования и анализа экспериментов. Она состоит из экспертной системы (БЗ, БД, блок логического вывода, блок объяснений), банка теорем, генератора планов, толкового словаря, тезауруса, блока игровых имитаций, пакета прикладных программ обработки данных. В концептуальном плане экспертная система состоит из четырех подсистем: постановка задачи и формализация объекта исследования, планирование эксперимента, статистический анализ, интерпретация. В случае экспериментального эксперимента подключается пятая подсистема, в которой реализована градиентная процедура для поиска оптимума.

При разработке экспертной системы использован язык Турбо Пролог. Технические средства реализации системы - IBM-совместимые персональные ЭВМ с операционной системой *MSDOS*. Рассмотрим некоторые вспомогательные компоненты.

Библиографическая информационная система написана на языке СУБД *Clipper*. В базе знаний хранятся ссылки на номера соответствующих рефератов. Эти рефераты можно просматривать из основной диалоговой программы, а можно воспользоваться всеми средствами баз данных, выйдя в операционную систему. Первое возможно, так как известна структура файлов *database*, а для обеспечения второго написана специальная программа.

Помощь и обучение являются существенными чертами системы и состоят из текущей помощи, тезауруса и словаря. Текущая помощь состоит из системной подсказки на нижней строчке экрана и из объяснения каждого вопроса, находящегося в специальном окне.

Тезаурус содержит терминологию по планированию эксперимента и описание его концептуальной схемы. С его помощью удобно знакомиться со структурой знаний системы и обучать

пользователя методам планирования эксперимента. Он написан на языке Турбо Пролог. Войти в него можно из соответствующего меню системы или из блока объяснений. К нему привязываются все остальные информационные системы ПЛАНЭКСа.

Толковый словарь знакомит пользователя с понятийным аппаратом, стандартизует терминологию во время диалога, несет справочную и обучающую функции. Словарь создан с помощью резидентной базы данных. Перед началом диалога в память загружается программа, которая управляет доступом к словарю, после чего вызывается основная диалоговая программа. В любой ситуации Вы можете, нажав определенный набор ключей, обратиться к словарю и по другому ключу вернуться к работе в прерванном месте. Вместе со словарем планируется создать учебник с аналогичной управляющей программой. Его объем может превысить 1 мегабайт текстов без сокращения рабочей памяти.

В системе используются программы для генерации планов эксперимента. Они написаны на языке С и хранятся на диске в отдельных файлах, откуда вызываются по мере необходимости. Для первичной статистической обработки данных и простых методов дисперсионного анализа используется пакет. Кроме этого имеется доступ к комплексу отечественных программ для планирования и анализа сложных многофакторных планов с ограничением на рандомизацию. Для оформления результатов работы предлагается использовать интегрированный пакет Фреймворк. Обращение с этими двумя пакетами требует знакомства с их описаниями и некоторой практики, зато такие знания имеют самостоятельную ценность для практической работы.

Л и т е р а т у р а

1. Маркова Е.В., Чеботарев Н.Г. АРМ специалиста по планированию эксперимента в экспертной системе ПЛАНЭКС, в сб.: I Всесоюзная школа-семинар "Разработка и внедрение в народное хозяйство персональных ЭВМ", тезисы докладов, часть V, ПОВТ, Минск, 1988, с. 134-137.
2. Маркова Е.В., Чеботарев Н.Г. Методо-ориентированная экспертная система по планированию и анализу эксперимента ПЛАНЭКС. Приборы и системы управления, №1, 1989г.

ОДИН ИЗ ПОДХОДОВ К СОЗДАНИЮ ЭКСПЕРТНОЙ СИСТЕМЫ ДЛЯ СТАТИСТИЧЕСКОГО КОНСУЛЬТИРОВАНИЯ

М.Д.Степанова, Т.И.Колесник

Рассматриваются принципы построения экспертной системы статистического консультирования (ЭССК). Данная система представляет собой один из возможных вариантов экспертной поддержки для статистических программных средств.

Назначение ЭССК – по описанию задачи пользователя определить необходимые для решения данной задачи статистические методы и программные средства. ЭССК должна обеспечить выбор соответствующих методов анализа, получение рекомендаций по их применению в условиях, описанных пользователем; обеспечить работу с непрофессиональным пользователем, анализируя его действия в процессе диалога; обеспечить работу со специалистом-экспертом, предоставляя ему автоматизированные средства представления знаний.

ЭССК включает в себя следующие компоненты: базу знаний (БЗ), систему логического вывода (интерпретатор правил продукции), систему объяснения действий экспертной системы, диалоговый интерфейс с пользователем, базу данных, систему поддержки базы данных и знаний.

База знаний содержит экспертные знания об объектах исследования: множество списков ограничений на методы, список ограничений на модели данных, набор элементарных правил типа продукций. Для представления знаний в БЗ используются два подмножества правил продукций и одно множество фреймов.

Первое подмножество правил продукций формирует знания экспертов об условиях применимости статистических методов, второе – знания о последствиях вероятностной модели данных.

Каждый из фреймов содержит информацию, касающуюся дополнительных аспектов знаний экспертов о методах анализа данных, критериях проверки гипотез и т.д.

Структура фрейма состоит из ряда слотов. Например, в первом из них может быть представлена информация об условиях, для которых применяется данная статистическая процедура (одна

или две анализируемых выборки, эксперимент с контрольной группой и т.п.), и предположениях, выдвигаемых при проверке гипотез (равенство объемов выборок и дисперсии, односторонняя или двусторонняя альтернатива и т.п.). Второй слот можно заполнить информацией, характеризующей преимущества или недостатки используемых процедур. Пользователь может запросить информацию, хранящуюся в слотах любого из фреймов, определяя имя необходимой статистической процедуры. Эта возможность реализуется с помощью модулей обработки фреймов.

Управляющая стратегия вывода строится как поиск на семантической сети следующим образом. Выбирается список ограничений, основанный на модели анализируемых данных. Проверяется справедливость каждого из ограничений (по взаимодействию с пользователем или путем проверки статистических гипотез). Если ограничения верны, начальный список гипотез уменьшается отбрасыванием методов, хранящихся в последующих элементах списка. Если ограничения ложны, отбрасываются методы в текущем элементе списка. Если справедливость ограничений не может быть подтверждена, система не отбрасывает ни один метод, и ответ дает пользователь.

База данных содержит список гипотез, описания статистических методов и моделей, описание программных средств и их дифференциальные признаки.

В качестве инструментальных средств для реализации системы используются персональный компьютер типа ЕС1840, операционная система MSDOS, базовые языки программирования Паскаль и Пролог.

СИСТЕМА АНАЛИЗА НЕЧИСЛОВОЙ ИНФОРМАЦИИ "САНИ"**С.П.Адамов, М.К.Вегнер, А.А.Халилеев, А.Г.Шимельмиц**

Система анализа нечисловой информации "САНИ" предназначена для обработки данных, измеренных в разнородных шкалах. В первую очередь она ориентирована на социально-экономические задачи, но может использоваться в биологии, медицине и других областях, где приходится иметь дело с нечисловыми данными. В ней реализованы многие классические и некоторые оригинальные процедуры. Особое внимание уделено методам анализа данных, изменяющихся во времени.

Система САНИ реализована на совместимых с IBM персональных компьютерах. Она требует около 1 М памяти на твердом диске, MS DOS не ниже версии 3.2 и графический адаптер типа CGA или более мощный.

Имеется развитая справочная служба: справки об используемом методе со ссылкой на литературу, или разъяснение сложившейся ситуации; справки об активных данных с некоторыми результатами проведенного анализа. Процедуры вызываются с помощью иерархического меню или непосредственно нажатием комбинаций двух клавиш.

РАБОТА С ДАННЫМИ

- Импорт данных из прямоугольных матриц "объект-признак" или "признак-объект", преобразование и сохранение данных в системном виде;
- Подвыборки: выбор или удаление объектов с зафиксированной комбинацией значений переменных.
- Новые переменные: задание категорий комбинациями значений имеющихся переменных; укрупнение категорий, разбиение количественных переменных на интервалы; всевозможные комбинации значений двух переменных (снижение размерности).
- Редактор.
- Экспорт данных в виде прямоугольных матриц "объект-признак" и "признак-объект".

-Представление больших массивов данных в оверлейном виде.

ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ

- Детерминационный анализ: дерево распределения объектов по категориям; определение комбинации, категорий для наилучшего предсказания.
- Графическое представление эволюционирующих данных в интерактивном режиме.
- Определение эталонных объектов.
- Анализ таблиц сопряженности: частоты; проценты; проценты по строкам и столбцам; отклонения от независимости; средние значения количественной переменной, средние разности двух количественных переменных в каждой клетке; мостеллеризация; сравнение двух таблиц; выбор формата; выделение экстремальных значений; меры парной связи; оригинальные алгоритмы выделения блочной и квази-блочной структур.
- Многомерные таблицы сопряженности.
- Анализ соответствий: одновременное представление объектов и категорий качественных переменных в евклидовом пространстве.

МОДЕЛИРОВАНИЕ.

- Гипотезы о таблицах сопряженности: симметричность; квази-симметричность; независимость; квази-независимость, маргинальная однородность.
- Дискретные случайные процессы: гипотезы однородности и марковости; характеристики эргодических и поглощающих марковских цепей; оригинальная процедура классификации, доставляющая максимум функции правдоподобия внутри групп.
- Динамическая модель квази-симметричности: эффекты привлекательности и отталкивания; матрица близости состояний.
- Лог-линейный анализ.
- Непараметрические критерии.

КОМПЛЕКСИРОВАНИЕ ПРОГРАММНЫХ СРЕДСТВ НА БАЗЕ СИСТЕМЫ СТАТИСТИЧЕСКОГО АНАЛИЗА

Х.Буш, Р.Полис, О.Штродах, А.Клейн

Система статистического анализа *SAS* в настоящее время является одной из самых развитых в мире /1/ и может служить хорошей основой для создания крупных программных "сверхкомплексов" /2/. В НПО "Силава" накоплен определенный опыт по обеспечению функционирования *SAS* в разных операционных системах ЭВМ ЕС в пакетном и диалоговом режиме, а также по взаимодействию с другими программными средствами.

Разработаны средства комплексирования:

- поддержка интерфейса *SAS-ABMDP* в среде ПДО, включая работу процедур *BMDP* и *CONVERT* на ПДО-файлах;
- интерфейс *SAS-СПЕКТР*, обеспечивающий чтение и запись из *СПЕКТР* -баз данных в *SAS* -файлы и наоборот;
- интерфейс с панельной системой *DMS/CMS* ;
- обеспечение работы с некоторыми процедурами предыдущих версий, в т.ч. *SAS/ETS* ;
- интерфейс с системой графического вывода *ГРАФОР*;
- поддержка сортировки и функций перевода регистров по кириллице, применение сортировки *SORTF* ;
- средства работы с библиотеками форматов как в виде *TXTLIB*, так и в виде ОС-библиотек;
- обеспечение сегментной загрузки и моделирования файлов в виртуальной памяти.

Проводятся работы по адаптации версии для персональных компьютеров.

Л и т е р а т у р а

1. Сильвестров Д.С. Программное обеспечение прикладной статистики. -М.: Финансы и статистика, 1988. -240 с.
2. Буш Х.К. Формирование инфраструктуры статистического анализа данных для отраслевой АСУ. -В кн.: Применение многомерного статистического анализа в экономике и оценке качества продукции. -Тарту, 1985. с.156-158.

СИСТЕМНАЯ ИНТЕГРАЦИЯ ПРИКЛАДНЫХ ПРОГРАММ НА БАЗЕ РАСШИРЕННОЙ КОНЦЕПЦИИ ГИПЕРТЕКСТА.

С. А. Егоров

Современные программные продукты обретают потребительскую ценность лишь в сочетании с информацией, определяющей деятельность людей по их использованию. Такая информация содержится в различных описаниях, справочниках, методических материалах, руководствах по эксплуатации и т.д. По материалам [1] при сравнительном анализе трех наиболее популярных статистических пакетов были приняты во внимание следующие критерии: простота использования, легкость изучения, полнота и качество документации, точность вычислений, соответствие стандарту, скорость вычислений. Приходится констатировать, что половина критериев оценивает пакет с точки зрения человеческой деятельности, другая половина оценивает машинный аспект. Для достижения высокого качества программного продукта необходимо обеспечить органическое единство всех средств поддержки деятельности как машины, так и человека. В работе с единых позиций рассмотрена деятельность человека при создании и эксплуатации интегрированной компьютерной системы. Показано, что системособразующим фактором может стать новый, предложенный автором, информационный объект - интерактивный гипертекст.

Гипертекст (ГТ) - это способ организации и представления электронного текста [2]. Под крупной рубрикой, связанной с подчиненными рубриками или подтемами, помещается тематически релевантная информация, доступ к которой осуществляется в соответствии с потребностями пользователя. Таким образом ГТ позволяет объединять информационные единицы с помощью ассоциативных связей и предоставляет средства для нелинейной последовательности чтения - обеспечивает более легкий доступ к информации.

Расширим концепцию ГТ, допустив возможность не только считывать, но и записывать информацию, а также возможность устанавливать ассоциативные связи с функциональными объектами. Тогда ГТ станет также удобным средством для внесения заметок пользователя, комментирующих основную

информацию, и средством создания "индивидуальных интерпретаторов", позволяющих интегрировать в единое целое как информационную, так и вычислительную модели системы. Информационная модель системы задается тройкой $S = (O, P, R)$, где O - множество объектов предметной области, P - множество свойств объектов, $R = O \times P$ - множество отношений между объектами. В реализации гипертекста, обсуждаемой в докладе, выделены следующие отношения R : синонимы, перекрестные ссылки (типа "см. также"), означающий-означаемый, обобщение-конкретизация, часть-целое, род-вид. Вычислительная модель диалоговой системы представляется схемой диалога - диаграммой переходов и состояний [3]. В докладе подчеркивается, что использование интерактивного ГТ в качестве системного интегратора позволяет удовлетворить многие требования к интерфейсу пользователя, в частности реализовать принцип "работать и учиться". Показано, что в смысле организации управления пакетом интерактивный ГТ покрывает многие требования проектировщиков диалоговых систем, позволяя единообразными средствами создать всевозможные меню (столбец, строка, матрица, разряженная матрица), а также создавать крупноформатные электронные бланки. В докладе представлен разработанный автором инструментальный комплекс программных средств организации и поддержки интерактивного ГТ.

Литература.

- Evans G.E., Newman A. "A Comparison of SPSS PC+, SAS PC and BMDP PC", Coll. Microcomput., 1988, 6, N2, 97-106, 122.
- Smith K.E. "Hypertext - linking to the Future", Online, 1988, 12, N2, 32-40.
- Егоров С.А., Стенчиков Г.Л., Паденко В.П. "Графический интерфейс пользователя для обработки результатов вычислительного эксперимента с моделью климата", в кн. Кибернетика и вычислительная техника, 3, стр. 287-295, 1987, М. Наука.

СИСТЕМА ИНТЕРАКТИВНОГО АНАЛИЗА МНОГОМЕРНЫХ ДАННЫХ (СИАМД)

И.С.Енюков, Н.Д.Вылегжанин, А.С.Соломоник

СИАМД предназначена анализа многомерных данных в интерактивном режиме на персональных компьютерах IBM PC/XT/AT и совместимых с ними. Отличительной характеристикой системы является широкое использование средств графического анализа, позволяющих проводить визуальную оценку структур многомерных данных. В частности, с этой целью используются методы целенаправленного проецирования [1]. Средства графического анализа представляют возможность использовать мультиоконную графику с подвижными окнами, проводить маркировку и удаление элементов в интерактивном режиме, осуществлять просмотр данных в режиме динамической графики (двухмерные и трехмерные вращения).

Управление системой осуществляется с помощью меню и функциональной клавиатуры. Существует развитая служба оперативной помощи пользователю ("HELP"-система).

Система контролирует действия пользователя, запрещая или выдавая соответствующие предупреждения в случае бессмысленных команд.

В настоящей версии СИАМД реализованы различные методы обработки данных, объединенные в следующие разделы:

- разведочный анализ данных для выявления структур многомерных данных на основе целенаправленного проецирования (кластеров, выбросов, нелинейных поверхностей и т.д.) и графического отображения этих структур в пространстве низкой размерности;
- многомерное шкалирование и нелинейное отображение данных;

- главные компоненты и факторный анализ;
- регрессионный анализ (регрессия с помощью целенаправленного проецирования, обычная линейная регрессия, регрессия с регуляризацией, пошаговая регрессия, робастная регрессия, непараметрическая пошаговая регрессия, преобразования линеаризирующие зависимость);
- дискриминантный анализ (канонический дискриминантный анализ, дискриминантный анализ на основе нескольких линейных комбинаций, метод опорной группы, метод k-ближайших соседей, пошаговый дискриминантный анализ);
- анализ соответствий;
- множественный анализ соответствий;
- кластер-анализ (метод k-средних, динамический градиентный кластер-анализ, иерархический кластер анализ).

СНАМД имеет средства ведения и организации данных пользователя. Возможен режим многократной обработки данных пользователя с изменением различных параметров, управляющих анализом (метрик, состава и преобразований переменных, состава объектов и т.д.), режим сравнения и результатов.

Л и т е р а т у р а

1. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Классификация и сокращение размерностей. М.: Финансы и статистика, 1989

СТАТИСТИЧЕСКАЯ СИСТЕМА "ЭКОНОМЕТРИКА"

Д.М.Кабанов, М.А.Седова, Г.В.Срагович

Система "Эконометрика" (название условное), представляет собой программное средство предназначенное для построения и анализа на персональных компьютерах типа IBM PC XT, AT экономико-математических моделей, основанных на регрессионных уравнениях. Система работает в интегрированной среде "Мастер".

Состав системы:

- блок управления данными, позволяющий осуществлять ввод и вывод информации, ее хранение в удобной для пользователя форме, редактирование, математические преобразования, обработку пропущенных значений.
- блок предварительной обработки, обеспечивающий вычисление дескриптивных статистик и проверку статистических гипотез.
- блок корреляционного анализа, содержащий набор процедур оценивания корреляционных зависимостей в различных режимах обработки пропусков.
- блок кластер-анализа (метода динамических сгущений и др.).
- блок регрессионного анализа, ориентированный на построение элементарных регрессионных моделей и содержащий методы робастного оценивания, конфлюэнтного анализа, оценивания в условиях мультиколлинеарности и гетероскедастичности.
- блок оценивания систем одновременных уравнений.
- блок ведения базы моделей, позволяющий осуществлять расчеты по созданным моделям.

Система организована таким образом, что с ней могут работать пользователи, не являющиеся ни профессиональными программистами, ни специалистами статистиками.

ОБ ОПЫТЕ РАЗРАБОТКИ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

А.В.Калита

Рассматриваются вопросы реализации программного обеспечения анализа временных рядов параметрическими методами для пакета СИГМАСТАТ (построение модели временного ряда; прогнозирование на основе имеющейся модели; коррекция прогнозов и параметров модели при поступлении новых отчетов; оценивание спектра на основании модели [1]).

Задача построения модели является основополагающей и в то же время трудно формализуемой, от которой зависит качество итоговых результатов. Принципиальными шагами при решении этой задачи является выбор подкласса моделей из общего их класса авторегрессии-скользящего среднего (АРСС), а также подбор параметров модели. Трудность формализации этих процессов и ориентация пакета на широкий круг пользователей наложила особенности на реализацию программного обеспечения. Предложено три режима функционирования пакета при построении модели АРСС [2].

Первый режим взаимодействия ориентирован на пользователей, хорошо знающих предметную область и разбирающихся в методах решения задач. При таком режиме инициатива принадлежит пользователю. Он может провести алгебраическое преобразование данных с помощью представленного "меню"; выбрать необходимую степень разности для обеспечения стационарности, используя автокорреляцию (АКФ) и частную автокорреляционную функции (ЧАКФ), отображаемые в виде графиков на дисплее; определить порядки авторегрессии и скользящего среднего и др. Для поддержки принятия решений пользователю (по требованию) могут быть представлены особенности поведения теоретических

АКФ и ЧАКФ для всех процессов. Кроме того, предоставляется возможность выбора критериев оценивания порядков модели; коэффициентов модели; методов диагностической проверки.

Второй режим взаимодействия ориентирован на пользователей-прикладников. В этом случае пакет полностью управляет процессом вычислений в автоматическом режиме, сам выбирает и оценивает модель временного ряда из заданного класса моделей.

Третьим режимом является режим "готового сценария", при котором пакет оценивает конкретную модель, заданную пользователем.

При решении любой задачи параметрического анализа временных рядов в рамках пакета пользователю может представляться, по его желанию, вспомогательная информация рекомендательного характера; возможность подготовки отчета по решаемой задаче; сохранение оцененных моделей для дальнейшего их использования.

Излагается опыт использования инструментальных средств системы СИГМАСТАТ [3] при разработке программных и информационных компонентов, обеспечивающих решение задач данного класса.

Л И Т Е Р А Т У Р А

1. Дж.Бокс, Г.Дженкис. Анализ временных рядов. Прогноз и управление. М., Мир, 1974.
2. О.В.Верева, А.В.Калита. Пакет прикладных программ СИГМАСТАТ: обработка и анализ временных рядов в диалоговом режиме //3 Всесоюзная школа-семинар "Программно-алгоритмическое обеспечение прикладного многомерного статистического анализа": Тез.докл., окт.1987г., пос.Цахкадзор (Арм. ССР).- Москва: 1987.- 4.1.- с.43-44.
3. И.Н.Парасюк, И.В.Сергиенко. Пакеты программ анализа данных: технология разработки.- М.:Финансы и статистика, 1988.- 159 с.

СТАТЕС-4 СО СТОРОНЫ ПОЛЬЗОВАТЕЛЯ

М.Каролин, Т.Тийтс

Дается обзор системы СТАТЕС-4, описание функционального наполнения, описание начальных данных и возможностей их преобразования, описание языка управления системой и описание задания на выполнение программы.

Система СТАТЕС-4 содержит программы по следующим разделам статистики: корреляционный и одномерный регрессионный анализ, множественный регрессионный анализ, дисперсионный анализ, дискриминантный анализ, факторный анализ, анализ главных компонент и канонических корреляций, анализ эмпирических распределений, анализ нелинейной, робастной и проективной регрессий.

Для управления работой программы создан язык управления (ЯУ) системой. ЯУ состоит из предложений, предложение имеет вид:

Ключевое слово - фактическое значение параметров
Ключевым словом является четырехбуквенное сокращение русского текста. Например, предложение для Функции Регрессии Многих переменных имеет вид:

$$\begin{aligned} \text{ФУРМ} &= \text{МОД}1(2,3)(X1+X2) \times P1 + \text{SIN}(P2 \times X3), \\ &\text{МОД}2(1,1)\text{LN}(\text{РОСТ} \times \times 2 + P1 \times \times 2) \end{aligned}$$

Начальные данные задаются в виде прямоугольной таблицы, где каждая строка принимается в качестве наблюдения, а каждый столбец - в качестве переменной. Предложением надо задать количество переменных (например, КОЛП=5). Переменные и наблюдения могут иметь наименования (НПЕР=ДА, ННАБ=ДА). Переменным можно присвоить наименований и во время работы программы (например, НАИМ=(X7-X3)(РОСТ,ВЕС)). Некоторые значения переменных могут отсутствовать. В этом случае надо задать коды-числа, обозначающие отсутствующие значения (например, КОДЫ=X5(-1)). Начальные данные вводятся либо с перфокарт (с образов перфокарт) по формату (на-

пример, $\Phi OPM=(A4,6F2.0))$, либо с набора данных на диске, куда они заранее записаны программами ввода.

Для преобразования начальных данных имеются следующие возможности.

На уровне ЯУ возможны выбор переменных по перечню (например, $ПЕРЕ=X5-X7$ ЗАВП=РОСТ), преобразование и добавление новых переменных (например, $ПРЕО=X5=ВЕС/РОСТ, X6=I^2(X2.EQ.XI+84)X24=X24*2$), выбор наблюдений по перечню (например, $ПНАБ=I-10, 2I-30$) или по логическому условию (например, $ОГРА=X1*2+X2*2.LE.IO.AND.X3.GT.O$). В некоторых программах возможна группировка данных. Например, в программах дисперсионного анализа 4 уровня фактора можно задать арифметической прогрессией, начинающейся с I5 с шагом 3 $ОФАН=E=X7(4, I5,3)$.

На уровне подпрограммы пользователя можно реализовать дополнительные возможности вычисления новых переменных и выбора наблюдений по условию.

На уровне отдельной программы возможны транспонирование матрицы начальных данных, объединение наборов данных и сжатие информации о регрессионной зависимости.

Для решения конкретной задачи с помощью программы системы СТАТЕС-4 пользователю надо составить задание на выполнение. В задании надо задать имя программы ($P=$), начальные данные и соответствующие предложения ЯУ. При необходимости задаются имя входного набора ($I=$), имя выходного набора ($O=$) (выходным набором является, например, файл коэффициентов регрессии или файл факторных нагрузок) и имя вспомогательного набора ($H=$). Вспомогательным (входным) набором программы является выходной набор предыдущей по исполнению программы (например, файл коэффициентов регрессии для программы вычисления остатков регрессионной модели). Имена и диспозиции дисков, на которых наборы I, O и H, задаются параметрами $VI=$, $DI=$, $VO=$, $DO=$, $VH=$ и $DH=$. Кроме того, можно присвоить задаче имя ($T=$), задать выходной класс печати результатов ($C=$), максимальное время для выполнения программы ($TM=$) и объем раздела памяти ($R=$). Если некоторый параметр опущен, то ему присваивается значение по умолчанию.

О НЕКОТОРЫХ ДОПОЛНИТЕЛЬНЫХ ВОЗМОЖНОСТЯХ СИСТЕМЫ СТАТЕС-4 ПО СРАВНЕНИЮ С ПАКЕТОМ ВМДР ПРИ АНАЛИЗЕ НЕЛИНЕЙНОЙ РЕГРЕССИИ И ЭМПИРИЧЕСКИХ РАСПРЕДЕЛЕНИЙ

М.Каролин, Т.Тийтс, В.Ольман

Одной из самых важных характеристик статистического пакета является многообразие моделей, которое можно подвергнуть анализу на уровне языка управления. В этом смысле язык управления СТАТЕС-4 обладает преимуществом перед языком ВМДР в области нелинейной регрессии и эмпирических распределений.

В системе СТАТЕС-4 имеется программа нелинейно-линейной регрессии одной переменной, в которой регрессионная зависимость между независимой и зависимой переменными задана в виде линейной комбинации базисных функций, зависящих нелинейно от неизвестных параметров. В этой программе количество оцениваемых параметров в подпрограмме минимизации уменьшается на число параметров линейной комбинации, что понижает размерность задачи минимизации и время решения.

В программах нелинейно-линейной и нелинейной регрессии одной переменной реализовано предварительное "сжатие" информации, что позволяет подвергать обработке большие массивы наблюдений, к которым не применимы программы ВМДР. "Сжатие" заключается в замене множества наблюдений на каждом интервале одним наблюдением, которое является значением либо линейного, либо квадратичного прогноза в среднем по интервалу значений.

В программах нелинейной, робастной и проективной регрессии одной переменной имеется возможность задать функцию регрессии по участкам, т.е. задаются границы участков

$G_0(P) = -\infty < G_1(P) < G_2(P) < \dots < G_{m-1}(P) < \infty = G_m(P)$
так, что на участке

$$[G_{i-1}(P), G_i(P)], \quad 1 \leq i \leq m,$$

функция P имеет вид

$$P = P_i(X, P),$$

где P - неизвестный оцениваемый K -мерный вектор-параметр.

В программах построения робастной и проективной регрессий задание функции невязки наблюдения и прогноза реали-

зовано на уровне языка управления, а не на уровне подпрограммы пользователя, как в пакете BMDP.

Предложение функции невязки в языке управления системой СТАТЕС-4 имеет вид:

$\text{ФУНЕ}=\text{КФ}(\text{Перечень переменных (КУ)} t_1, \dots, t_{kx-1}, t_k, \dots, t_{kx}, \dots), \dots$
где КФ – код функции регрессии,

КУ – количество участков, на которое разбита ось задания функции невязки,

t_1, \dots, t_{kx} – вещественные положительные числа в возрастающем порядке, определяющие участки,

y_1, \dots, y_{kx} – функции невязки на заданных участках.

В программах анализа эмпирических распределений по критерию χ^2 – Пирсона проверяется гипотеза о согласии эмпирического и теоретического распределений. Теоретическое распределение может быть выбрано из II библиотечных двупараметрических распределений или задано пользователем в виде функции распределения или плотности. Распределение можно задавать по участкам аналогично тому, как это делается для регрессионных кривых в нелинейной регрессии. Кроме того распределение может быть задано смесью распределений. При выполнении программы оцениваются параметры распределений и их весовые коэффициенты в смеси. Параметры теоретических распределений оцениваются по выбору пользователя либо по методу хи-квадрат, либо методом максимального правдоподобия.

Л и т е р а т у р а

1. В.Ольман, М.Каролин, М.Ломп, И.Петерсен, К.Пукк. Программы анализа эмпирических распределений и нелинейной регрессии. Таллинн, АН ЭССР, 1985.
2. В.Ольман, М.Каролин. Программы построения робастной и проективной регрессии. Таллинн, АН ЭССР, 1988.
3. Т.Тийтс, М.Каролин. СТАТЕС-4. Система программ прикладной статистики. Таллинн, АН ЭССР, 1988.
4. Dixon W.J. BMDP Biomedical Computer Programs. Berkeley, University of California Press, 1981.
5. Т.Тийтс, Э.Вийкманн, И.Каролин, К.Пукк. СТАТЕС-2. Средства управления и руководство по применению. Таллинн, АН ЭССР, 1986.

МОДИФИЦИРОВАННЫЙ "ПАКСТАТ"

Т. Кинкар

Пакет ПАКСТАТ для ПЭВМ является унифицированной библиотекой программ статистического анализа и вспомогательных программ, написанных на языке BASIC.

К преимуществам пакета ПАКСТАТ можно отнести следующие его свойства:

- * скромные требования к аппаратным средствам,
- * широкие возможности для модифицирования пакета, учитывая конкретные требования области применения,
- * наличие исходных текстов программ.

Учитывая особенности пакета, он является доступным и оперативным средством для первичной обработки небольшого количества данных используя при этом ЭВМ (например, Искра 1030.11) на которой работа с более мощными пакетами (например, Statgraf) затруднена.

При работе с пакетом выяснилось, что для более успешного его использования полезно добавить в первоначальный комплект некоторые процедуры и усовершенствовать часть уже имеющихся с помощью небольших изменений.

К примеру, дополнения были введены в процедуру CHISQ, которая вычисляет χ^2 статистики. К уже имеющимся возможностям (критерий Бартлетта, разности среди пропорций, равные и неравные эмпирические частоты) было добавлено сравнение процентов. Была обобщена также процедура вычисления ранговых корреляций для случая со связанными рангами.

Некоторые дополнения были введены и в систему обслуживания файлов и манипулирования данными.

Л и т е р а т у р а

1. Многофункциональная библиотека программ статистического анализа ПАКСТАТ. Руководство пользователя. Тарту, 1987. 104 стр.

ПАКЕТ СТАТИСТИЧЕСКИХ ПРОГРАММ "АНАЛИЗ ДАННЫХ"

В.О.Мазур, А.Ю.Юровский, А.М.Никифоров, Е.Г.Ардаева,
А.С.Титов, К.М.Богатов, Е.Е.Кожаринова, С.В.Углова,
С.В.Сливко-Кольчик

Применение статистических методов обработки данных в экономике, технических науках, биологии, медицине, психологии и т.п. ставит задачу создания программного продукта, с которым могли бы работать специалисты-нематематики и, с другой стороны, который был бы функционально насыщен. Предлагаемый пакет статистических программ "Анализ данных" пытается решить эти задачи. Он представляет собой диалоговую систему для обработки статистических данных в виде таблиц "объект-признак". Все программы пакета приспособлены для работы с таблицами, в которых есть пропущенные наблюдения.

Пакет состоит из трех разделов. Каждый раздел снабжен своим меню и инструкциями по применению методов, содержащихся в этом разделе.

Первый раздел - служебный. Он представляет собой экранный редактор таблиц типа "объект-признак" и предназначен для работы с данными (ввод, редактирование, просмотр на экране, манипулирование таблицами и т.д.).

Во втором разделе собраны методы визуализации данных. Это диаграммы рассеяния, двумерное шкалирование и ядерная оценка плотности распределения признака из таблицы данных. Все графики выводятся на экран и, по желанию, на печать. В программе двумерного шкалирования применен статистический подход, позволяющий существенно ускорить процесс вычисления. В программе построения ядерной оценки параметр сглаживания выбирается адаптивно, на основе данных.

Третий раздел представляет собой набор методов вероятностно-статистической обработки данных. Они реализованы в следующих программах:

- корреляционный анализ; вычисляются парные и частные выборочные коэффициенты линейной корреляции между признаками

таблицы данных, подчеркиваются значимо отличающиеся от нуля коэффициенты при общей, а также при односторонних альтернативах;

- регрессионный анализ (множественная линейная регрессия); вычисляются оценки коэффициентов регрессии, проверяются гипотезы о равенстве нулю каждого из коэффициентов, а также проводится дисперсионный анализ регрессии; для всех критериев вычисляется фактически достигаемый уровень значимости;
- факторный анализ; вычисляются собственные значения и векторы, определяются матрицы факторных нагрузок, когда факторы - главные компоненты и после вращения методом варимакс;
- проверка гипотез об однородности двух одномерных выборок : проверка гипотезы о совпадении математических ожиданий для выборок из двух нормальных распределений с произвольными дисперсиями (критерий Крамера-Уэлча), критерий разностей (одновыборочный критерий Стьюдента), обобщенный критерий Колмогорова-Смирнова для произвольных распределений (не обязательно непрерывных); для всех критериев вычисляется фактически достигаемый уровень значимости;
- анализ одномерных распределений; вычисляется набор из 13 одномерных статистических характеристик (оценки среднего, дисперсии и т.п.) и строится гистограмма и эмпирическая функция распределения для наблюдений из столбца таблицы данных.

Все результаты, получаемые в процессе работы с пакетом, окончательные : обращение к статистическим таблицам не требуется. В пакет включены как известные алгоритмы, так и некоторые новые результаты авторов.

Литература

1. Енюков И.С. Методы, алгоритмы, программы многомерного статистического анализа. М. "Финансы и статистика", 1986.

СИСТЕМА СИГМАСТАТ: ФУНКЦИОНАЛЬНЫЕ И ИНСТРУМЕНТАЛЬНО-ТЕХНОЛОГИЧЕСКИЕ ВОЗМОЖНОСТИ

И.Н.Парасюк, О.В.Верева, А.В.Калита,
Е.С.Карпинка, А.И.Провоторь, Н.П.Скопецкая

Система СИГМАСТАТ – дальнейшее развитие статистической системы ДЕЛЬТАСТАТ [1] в различных аспектах: в функциональном, в системном, в интеллектуальном и в инструментально-технологическом при сохранении базовых концептуальных моделей системы, языка общения и способов представления знаний, базовых унифицированных механизмов межпрограммного интерфейса, интеллектуального интерфейса и др.

В докладе конструктивно рассматриваются ключевые вопросы разработки системы СИГМАСТАТ в указанных аспектах.

В функциональном отношении система СИГМАСТАТ представляет собой проблемно-ориентированную программную систему для обработки данных различными методами как непараметрического спектрального анализа (можно проводить предварительную обработку временных рядов во временной области, оценивать статистики второго порядка, проводить развернутый взаимный спектральный анализ), так и параметрического спектрального анализа (модели Бокса-Дженкинса), методами регрессионного анализа и др.

Системное обеспечение – это многоуровневая семантически управляемая на основании знаний система типа "операционной оболочки" для поддержки компьютерной технологии решения задач обработки данных математическими методами. При этом процесс обработки данных рассматривается как многошаговый динамически-непрерывный процесс, который может носить как альтернативный (в основном на поисковых этапах), так и канонический (на этапе массовой обработки накопленного фонда

данных) характер.

В системе СИИМАСТАТ выделены подсистемы трех уровней:

Подсистема I -го уровня представляет собой специализированную мультипрограммную управляющую систему в режиме разделения времени. Её основное назначение – управление прикладной программной системой, одновременно с которой могут быть инициированы с помощью функциональных клавиш процессы сервисного и инструментального характера.

Подсистема 2-го уровня – суть семантически управляемый интерпретатор языка директив. Она управляет системными модулями, обеспечивает ведение памяти, ведение данными, межмодульный интерфейс, поддержку межсистемного интерфейса многошагового режима работы, режима многовариантного счета (автоматическая смена вариантов).

Подсистема 3-го уровня – это семантически-управляемый прекомпилятор интерпретирующего типа. Она поддерживает интеллектуальный интерфейс с пользователем на базе использования современных методов ведения диалога, поиск и синтез алгоритмов интегрированной обработки данных, межмодульный интерфейс и интерпретацию прикладных программ, участвующих в решении прикладных задач.

Для управления базами данных матричной структуры с имеющимися строками и столбцами используется специализированная СУБД.

Инструментально-технологические комплексы поддерживают разработку сложных и трудоемких процессов: программирование знаний, программирование сценариев диалогового взаимодействия, сборка программных систем из стандартизованного программного фонда, генерация семейства родственных систем с заданными функциональными возможностями и др.

Л И Т Е Р А Т У Р А

1. Парасюк И.Н., Сергиенко И.В. Пакеты программ анализа данных: технология разработки. – М: Финансы и статистика – 1988. –

ИНТЕГРИРОВАННАЯ СИСТЕМА ОБРАБОТКИ СТАТИСТИЧЕСКИХ ДАННЫХ - СТАТЕС - 4

Т.Тийтс, М.Каролин

Система СТАТЕС-4 разработана в Институте кибернетики АН ЭССР для ЕС ЭВМ с операционной системой ОС ЕС. В систему СТАТЕС-4 включены 56 программ [1-7], которые реализуют широкий спектр методов прикладной статистики. Эта интегрированная система [8] представляет собой одну трехобластную оверлейную программу из 80 сегментов, т.е. ППП многофункционального типа в соответствии с классификацией Сильвестрова [9].

По функциональному наполнению система СТАТЕС-4 наиболее развита в области дисперсионного (ковариационного) анализа [4] и нелинейной регрессии [6,7], где остальные оригинальные отечественные разработки отстают от зарубежных.

В системе СТАТЕС-4 реализован удобный для пользователя язык управления (ЯУ) при помощи четырехбуквенных ключевых слов [1,8], которым присваиваются фактические значения параметров. Фактические значения параметров представляются в виде, близком к естественному языку. Например, в отличие от пакета ППСА [10], где преобразование переменных весьма искусственное, пользователь системы СТАТЕС-4 задает просто аналитические формулы вычисления (новых) переменных. Благодаря возможности идентифицировать переменные при помощи наименований (это предусмотрено и в проекте пакета АСТА [11]), вместо обозначений переменных (буква с индексом) можно задавать их наименования, что выглядит еще проще.

У значительной части ключевых слов задание фактических значений параметров необязательно и в случае их отсутствия они принимают значения, принятые по умолчанию.

Для "стыковки" системы СТАТЕС-4 с базой данных пользователя имеются две возможности:

1) оформить результаты запроса с базы данных так, как этого требует система СТАТЕС-4,

2) составить программу USER (на фортране), которая

последовательно образует наблюдение из данных, записанных в базу данных пользователя.

Например, для использования данных из архива BMDP[12] достаточно написать 4 оператора в подпрограмму USER.

Выходная информация системы STATEC-4 оформлена на русском языке; она наглядна и понятна пользователю без дополнительной консультации.

Л и т е р а т у р а

1. Тийтс Т., Вийкманн Э., Каролин М., Пукк К. STATEC-2. Средства управления и руководство по применению. Таллинн, АН ЭССР, 1986.
2. Кукс Я., Вийкманн Э., Пукк К. Программы корреляционного и одномерного регрессионного анализов. Таллинн, АН ЭССР, 1980.
3. Кукс Я., Тийтс Т., Вийкманн Э. Программы множественного регрессионного анализа. Таллинн, АН ЭССР, 1979.
4. Кукс Я., Тийтс Т. Программы дисперсионного анализа. Таллинн, АН ЭССР, 1980.
5. Кукс Я., Каролин М., Тийтс Т. Программы дискриминантного и факторного анализов. Таллинн, АН ЭССР, 1981.
6. Ольман В., Каролин М., Ломп М., Петерсен И., Пукк К. Программы анализа эмпирических распределений и нелинейной регрессии. Таллинн, АН ЭССР, 1985.
7. Ольман В., Каролин М. Программы построения робастной и проективной регрессии. Таллинн, АН ЭССР, 1988.
8. Тийтс Т., Каролин М. Система программ прикладной статистики STATEC-4. Руководство по использованию (Дополнения к STATEC-2). Таллинн, АН ЭССР, 1988.
9. Сильвестров Д.С. Программное обеспечение прикладной статистики. М., Финансы и статистика, 1988.
10. Енюков И.С. Методы, алгоритмы, программы многомерного статистического анализа: Пакет ППСА.М., Финансы и статистика, 1986.
11. Кощеев В.А. Автоматизация статистического анализа данных: пакеты прикладных программ. М., Наука, 1988.
12. Dixon W.J. BMDP Biomedical Computer Programs. Berkeley, University of California Press, 1981.

ИНТЕРАКТИВНЫЙ РЕЖИМ КЛАССИФИКАЦИИ В СВМ.

Ю. Вилисмяз

В ТГУ уже 10 лет на ЭВМ ЕС-1060 в пакете СТЕЛЛА работает комплекс программ классификации данных, в том числе программа для создания таблицы сопряженности данных двух признаков и для вычисления некоторых важнейших характеристик их совместного распределения (т.н. двумерный статистический анализ) [1]. При этом обработка данных производится в пакетном режиме, при котором пользователь, однако, оперативно не может вмешиваться в ход вычислений. Кроме того, как показывает практика, пакетный режим способствует массовому применению этой методики, что часто приводит к малообоснованной затрате вычислительных ресурсов и печатной бумаги.

Чтобы избежать эти недостатки, программа двумерного анализа в СВМ была переведена в диалоговый режим, позволяющий пользователю самому управлять ходом работы программы и точно определить необходимое в задаче количество выводимой информации. Входными данными для этой программы являются стандартный файл данных пакета СТЕЛЛА на диске ОС ЕС и команды, введенные с клавиатуры дисплея. Анализ пары признаков разделяется на более типичные для практики блоки: создание частотной таблицы сопряженности и таблицы распределения; вычисление некоторых числовых характеристик совместного распределения; вычисление условных эмпирических распределений; образование регрессионных линий и корреляционного поля.

Все результаты выводятся только на экране дисплея.

В начале работы вводятся номера двух анализируемых признаков. Перед каждым блоком обработки спрашивается, надо ли выполнить рассматриваемый блок. Предусматривается возможность возврата к началу любого предыдущего блока и повторения его. Когда анализ заданной пары признаков окончен, задаются номера двух следующих признаков и работа продолжается.

Л и т е р а т у р а

1. Vilismae, J. Kahemõõtmeline statistiline analüüs. Statistilise andmetöötlussüsteemi metoodiline juhend, Programme kõigile, Tartu, 1984, 79-86.

ОБ ОДНОМ ПОДХОДЕ К ПРОГНОЗИРОВАНИЮ РЯДОВ ДИНАМИКИ

В.И.Елейко

Важной задачей многомерного статистического анализа является разработка методов прогнозирования уровней временных рядов небольших объемов [1].

В докладе рассматривается способ прогнозирования уровней рядов динамики, основанный на многократном построении трендов. Для заданного временного ряда методами наименьших квадратов и экспоненциального сглаживания строятся, например, последовательности линейных уравнений регрессий

$$y_t = a_0 + a_1 t, \quad \hat{y}_{t+\ell} = \hat{a}_0 + \hat{a}_1 \ell$$

и вычисляются прогнозные значения на один шаг \hat{y}_{t+1} . Здесь a_0 , a_1 и \hat{a}_0 , \hat{a}_1 - оценки параметров трендов, полученные методами наименьших квадратов и экспоненциального сглаживания; t - время; ℓ - период упреждения.

Далее, для временного ряда, образованного из разностей истинных значений ряда и соответствующих прогнозных значений вычисляется средняя квадратическая ошибка σ . Критерием выбора и его статистическим обоснованием полученного уравнения основной тенденции для прогнозирования является выполнение условия $\sigma < \hat{\sigma}$, где $\hat{\sigma}$ - оценка прогнозного значения \hat{y}_{n+1} . Если указанное условие не выполняется, то данную процедуру повторяют для построения других трендов.

С целью реализации предложенного подхода вычисления прогнозных значений на ЭВМ составлены на алгоритмическом языке ПАСКАЛЬ соответствующие программы для персональных компьютеров ЕС - 1840. Рассмотрен численный пример.

Л и т е р а т у р а

Г. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Исследование зависимостей. - М.: Финансы и статистика, 1985.

О БИБЛИОТЕКЕ ПРОГРАММ ПОСТРОЕНИЯ ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ ПРИ ОЦЕНИВАНИИ ВЕЛИЧИНЫ ЭФФЕКТА ВОЗДЕЙСТВИЯ

Е.И.Жилинская

Библиотека программ создана в процессе решения задачи оценивания эффективности противорадовой защиты по значениям метеорологических характеристик $X_k^{(j)}$.

Пусть $X_k^{(j)}$ ($k=1, \dots, n$, $j=1, 0$) независимые, случайные, по K равно распределенные величины, причем при $j=1$ измерения соответствуют экспериментальным единицам с воздействием, $j=0$ — контрольным. В качестве меры эффективности берутся отношения математических ожиданий: $d = EX^{(1)} / EX^{(0)}$, для которых требуются построить односторонние доверительные интервалы.

Библиотека программ, построения доверительных интервалов для величины эффекта воздействия разбивается на следующие разделы:

1. проверка согласия данных с типом распределения: двухпараметрическое Гамма и Вейбулла-Гнеденко, с помощью модификации критерия X^2 для непрерывных распределений.
2. построение асимптотических односторонних доверительных интервалов вида $(EX / EX^{(0)} - \varepsilon, \infty)$, при условии, что случайные величины $X_k^{(j)}$ имеют тип распределения Гамма / или Вейбулла-Гнеденко/, а в результате воздействия тип распределения не меняется, а меняется лишь значение двумерного параметра.
3. Построение доверительных интервалов методом бутстреп, не предполагающим знания типа распределения данных.

Программы использовались при оценивании эффективности противорадового эксперимента на Северном Кавказе, в качестве входных данных брался ряд метеорологических характеристик, в частности, максимальная величина фактического диаметра града, выпавшего на засеваемой и контрольной площадях; был найден положительный эффект воздействия.

Литература.

1. Е.И.Жилинская, Э.М.Кудлаев, статистический алгоритм построения доверительных интервалов для величины эффекта воздействия в погодных экспериментах, тезисы докл. III всесоюз. школы-семинара, Цахкадзор, М.1987, с.296-297.

БЫСТРЫЙ МЕТОД ОЦЕНИВАНИЯ ПОЛИНОМИАЛЬНОГО ТРЕНДА ВРЕМЕННОГО РЯДА

В.Йоала

Для оценивания коэффициентов полиномиального тренда временного ряда обычно применяется метод наименьших квадратов (МНК), который среди линейных методов оценивания является наилучшим несмещенным методом. Недостатком этого метода является большое количество умножений и сложений при оценивании, что приводит к относительно медленным программам на ЭВМ.

В докладе предлагается быстрый метод оценивания, который при незначительной потере в точности по отношению к МНК позволяет существенно сократить продолжительность оценивания на ЭВМ. При этом существуют два подхода по реализации этого метода - применение разностного оператора и использование перемножаемых прямоугольных периодических функций.

Несмотря на ограничения на объем выборки, предлагаемый метод может успешно заменить МНК в тех случаях, когда ряды длинные, а также при планировании эксперимента. Приводятся результаты вычислительных экспериментов.

Л и т е р а т у р а

1. Йоала В. Оценивание коэффициентов полиномиального тренда временных рядов с применением разностного оператора. // Труды ВЦ Тартуского государственного университета. - 1988. - № 56. - С.3-16.
2. Joala V. An Estimation Method for Coefficients of Polynomial Trend in Time Series // Acta et Commentationes Universitatis Tartuenssis. - 1988. - N 798 : Statistical and Probabilistic Models. - p. 68-75.

ЭКСПЕРТНАЯ ПОДДЕРЖКА ВЫБОРА МОДЕЛИ БОКСА - ДЖЕНКИНСА В СИСТЕМЕ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ МЕЗОЗАВР

С.Е. Кузнецов, Д.В. Левандо

В работе описывается модуль автоматического подбора модели Бокса-Дженкинса в системе статистического анализа временных рядов МЕЗОЗАВР. Модель Бокса-Дженкинса обладает хорошими аппроксимационными свойствами и эффективна для построения краткосрочных прогнозов. Структура модели описывается четырьмя (а в сезонном варианте - восемью) целочисленными параметрами. Этап определения порядка модели представляет собой трудноформализуемую задачу, решение которой в состоянии дать лишь квалифицированный специалист в области временных рядов. Переборные методы здесь крайне неэффективны из-за трудоемких вычислений. Кроме того, оценить качество полученной модели может только специалист. Поэтому в настоящее время эта модель применяется достаточно редко.

Подбор модели Бокса-Дженкинса сильно зависит от опыта и субъективного мнения эксперта. Для пользователя, не являющегося специалистом в области временных рядов, модель представляет интерес лишь в том случае, когда можно получать результаты, не обращаясь к эксперту. С целью сделать использование модели Бокса-Дженкинса доступным не только специалистам, в системе МЕЗОЗАВР предусмотрена возможность выбора модели в автоматическом режиме (как для несезонных, так и для сезонных временных рядов). На первом шаге определяется необходимый порядок разности и целесообразность включения константы, а также находятся стартовые модели, с которых начинается поиск. Дерево поиска учитывает результаты расчетов на каждом шаге. Переходы между состояниями отражают действия эксперта при подборе модели. Рассмотренные модели классифицируются на четыре группы: "хорошие", "недоопределенные" (т.е. предположительно, необходимо включение дополнительных параметров), "переопределенные" (предположительно, содержат лишние параметры) и "непонятные". Для оценки качества модели принимаются во внимание остаточные автокорреляции и статистические характеристики параметров мо-

дели, объединенные в пять (а в сезонном случае – в восемь) различных характеристик качества. Используется аппарат нечетких функций принадлежности. Окончательный выбор модели представляется пользователю.

Диапазон поиска ограничен – для несезонных моделей количество членов авторегрессии варьируется в пределах от 0 до 8, а количество членов скользящего среднего – до 4. В свою очередь, для сезонных моделей – простая авторегрессия – до 3, сезонная авторегрессия – до 2, скользящие средние – до 2, а сезонные скользящие средние не превосходят 1. Визуально процесс перебора представлен в виде обхода таблицы, в которую заносятся результаты.

Пользователь может в любой момент прервать процесс автоматического поиска и перейти в интерактивный режим, в котором он может получить справку о любой из уже рассмотренных моделей. В этой справке содержится обоснование экспертного заключения о модели. Кроме того, пользователь может вызвать для расчетов любую другую модель в указанных диапазонах. В том случае, когда с точки зрения системы проведение расчетов не имеет смысла, на экране появится соответствующее предупреждение. Пользователь имеет возможность получить полную информацию о любой модели, с тем чтобы составить собственное заключение о ее качестве. Ему доступны также значения критерия Акайке. Предлагаемые программой решения в бесспорных случаях совпадают с выбором специалиста, а в сомнительных – являются одной из возможных точек зрения.

СТАТИСТИЧЕСКАЯ СИСТЕМА ПО АНАЛИЗУ ВРЕМЕННЫХ РЯДОВ "МЕЗОЗАВР"

С.Е.Кузнецов, А.А.Халилеев

Система МЕЗОЗАВР разработана для персональных компьютеров класса IBM PC/XT/AT и предназначена для анализа временных рядов умеренной (до нескольких тысяч наблюдений) длины. Она ориентирована на специалиста средней квалификации, имеющего достаточное представление о методах анализа временных рядов.

Существенным отличием системы МЕЗОЗАВР от большинства других систем является то, что в ней каждому значению временного ряда автоматически сопоставляется соответствующий момент времени, что учитывается при осуществлении любых операций над временными рядами, построении графиков, построении моделей и др. Система может работать с временными рядами с периодичностью от 1 минуты до нескольких лет. Возможно рассмотрение рядов и в условной временной шкале.

Результаты расчетов по желанию пользователя могут быть сохранены как новые переменные в оперативной памяти (и впоследствии записаны в файл). Системе известна природа всех вновь создаваемых переменных.

Поскольку анализ графиков временных рядов является необходимым этапом исследования, система МЕЗОЗАВР включает в себя развитую интерактивную графическую поддержку.

Управление системой осуществляется с помощью меню и функциональной клавиатуры. Существует служба "HELP".

Система контролирует действия пользователя, запрещая или выдавая соответствующие предупреждения в случае бессмысленных команд, например, арифметических операций над рядами с несовместимой временной сеткой, выделения тренда из автокорреляционной функции и т.п.

Функциональное наполнение системы МЕЗОЗАВР:

- ввод-вывод и предварительная обработка данных;
- работа со стандартными файлами системы;

- "импорт" и "экспорт" файлов, записанных в формате ASCII и dBASE;
- редактор данных;
- графический редактор данных;
- преобразования временных рядов (интерпретатор формул над рядами и числами);
- интерактивный графический анализ:
 - график временных рядов (до 10 рядов одновременно, возможность увеличения фрагментов графика, несколько типов шкал и другие возможности);
 - гистограмма;
 - корреляционное поле;
- интерактивные процедуры сглаживания;
- автокорреляционная функция (простая и частная);
- периодограмма;
- спектр и кросс-спектр;
- корреляционная матрица;
- кросс-корреляционная функция;
- заполнение пропущенных наблюдений;
- обнаружение выбросов и разрывов;
- статистические тесты;
- скользящее среднее;
- экспоненциальное сглаживание;
- частотная фильтрация;
- параметрическое оценивание линейных и нелинейных моделей тренда;
- выделение тренда с учетом сезонности;
- авторегрессионная модель;
- модель Бокса-Дженкинса с экспертной поддержкой;
- множественная линейная регрессия;
- робастная линейная регрессия;
- пошаговая линейная регрессия;
- архив моделей;
- справочная служба по терминологии и применяемым методам.

ПАКЕТ ПРОГРАММ "ЛОТР" ДЛЯ РЕШЕНИЯ ЗАДАЧИ ГРУППИРОВКИ ОБЪЕКТОВ

Т.М.Пестунова

Пакет программ "Лотр" предназначен для решения задачи группировки объектов, являющейся одной из наиболее распространенных в анализе данных. Необходимость в её решении возникает также при разработке систем распознавания, прогнозирования, оценки качества продукции. Исходная информация представляется в виде таблицы "объект-признак".

Основу пакета составляют программы группировки, разработанные в рамках подхода с использованием логических функций [1,2], который ориентирован прежде всего на обработку таблиц, обладающих рядом свойств, которые препятствуют эффективному применению других методов:

- высокая априорная неопределённость данных;
- наличие признаков, измеренных в разных шкалах;
- наличие в таблице пропусков отдельных значений;
- число объектов сравнимо с числом признаков.

Принципиальные особенности подхода следующие.

1. При обработке разнотипных данных не требуется преобразования признаков к одному типу, что обеспечивает наиболее полное использование исходной информации.

2. При наличии пропуска соответствующие объект и признак не исключаются, в ходе группировки возможно приближённое его прогнозирование с точностью до некоторого множества значений.

3. Наряду с выделением групп данных происходит формирование соответствующих каждой группе закономерностей в признаковом пространстве на языке логических функций конъюнктивного и дизъюнктивного вида, который в силу своей простоты и наглядности облегчает содержательную интерпретацию и дальнейшее использование результатов.

4. В процессе группировки для каждой выделенной группы происходит исключение неинформативных признаков, причём эти подмножества признаков могут быть различными для разных групп.

5. Не используются меры близости и расстояния в пространстве признаков, введение которых во многих случаях связано с методологическими трудностями.

6. Разработанный аппарат логических функций может использоваться для построения комбинированных алгоритмов, сочетающих преимущества данного подхода с достоинствами других методов.

Пакет включает блоки предварительной и основной обработки. Блок предварительной обработки предназначен для получения дополнительной информации об исследуемых данных, которая необходима для реализации основных алгоритмов или представляет самостоятельный интерес для пользователя. В него включены программы, осуществляющие анализ пропусков, расстановку грааций, вычисление некоторых общих характеристик таблицы и др. Блок основной обработки дает возможность получения различных типов группировок.

1. Группировка типа "центров сгущения" при одновременном возможном присутствии неразгруппированных объектов. Требования к непересекаемости групп определяются пользователем.

2. Группировка типа дихотомического дерева.

3. Группировка, разделяющая сложные структуры данных на таблицах, не имеющих пропусков (комбинированный алгоритм, основанный на логических функциях и идеях работы [3])

Программы пакета успешно применялись для решения ряда практических задач медицинской диагностики, геологического прогнозирования, классификации лесных пород.

Л и т е р а т у р а

1. Лбов Г.С., Пестунова Т.М. Группировка объектов в пространстве разнотипных признаков. - В кн.: Анализ нечисловой информации в социологических исследованиях - М., Наука, 1985.

2. Лбов Г.С., Пестунова Т.М. Построение дерева разбиений в задаче группировки объектов с использованием логических функций. - В кн.: Вычислительные системы, вып. II7, Новосибирск, 1986.

3. *Magowan, B. Shomura. A Non-parametric Algorithm for Detecting clusters Using Hierarchical Structure - IEEE Trans on Pattern Analysis and Machine Intelligence Vol 2, No 4, 1980, pp 392-399*

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ АНАЛИЗА ПАРНЫХ СРАВНЕНИЙ

Т.А.Пригарина

Преимущество парных сравнений (п.с.) в прикладных исследованиях предпочтительности объектов (их важности, качества и т.д.) заключается в возможности более точного описания индивидуальных предпочтений по сравнению с обычным ранжированием или балльными оценками. Это оказывается важным в экспертизах и пилотажных опросах, позволяя, в частности, повышать надежность их результатов. Вместе с тем п.с. трудоемки и для эксперта (респондента) и для исследователя, а методы их обработки рассеяны по множеству публикаций. Все это делает настоятельную разработку пакета программ с элементами экспертной системы, обеспечивающую широкое применение п.с. в практике опросов.

Основные черты первой версии проектируемого пакета следующие. Входными данными являются матрицы п.с., элементы которых кодируют для каждой пары объектов ответы трех типов: "лучше", "хуже", "одинаково". Несравнимость по предпочтению в первой версии не допускается, хотя учитывается возможность случайных пропусков. Так как п.с. предназначены прежде всего для задач с неопределенным или многомерным критерием предпочтительности, когда индивидуальные предпочтения могут заметно отклоняться от одномерной шкалы, порядковой и тем более количественной, обработка матриц с элементами, отражающими интенсивность предпочтения на парах, пока не предусматривается. В принципе большинство используемых методов обобщается и на этот случай.

Структура пакета соответствует трем основным задачам анализа п.с., которые рассматриваются как последовательные этапы обработки. Выделены разделы проверки информативности (одномерности) предпочтений, оценки их согласованности и агрегирования – получения общего упорядочения или количест-

венных оценок. Так как в опросах с большим числом объектов важно снизить нагрузку на эксперта, предусмотрен раздел планирования эксперимента. Система позволяет пройти все этапы обработки без выхода из программы, то есть осуществить полную обработку введенных данных. Чтобы помочь пользователю сориентироваться в имеющихся возможностях и вариантах, предполагается организовать диалог с подсказками и комментариями.

Пакет находится в стадии алгоритмической разработки. В последующих версиях планируется включение в обработку неполных данных (матриц п.с. с системными пропусками), а также решение нетрадиционных задач анализа п.с.

Используемые методы описаны в работах [1 - 5].

Л и т е р а т у р а

1. Дэвид Г. Метод парных сравнений. М., Статистика, 1978.
2. Шмерлинг Д.С., Дубровский С.А., Аржанова Т.Д., Френкель А.А. Экспертные оценки. Методы и применение.// Статистические методы анализа экспертных оценок. М., Наука, 1977.
3. Пригарина Т.А., Поляков В.В. Анализ информативности индивидуальных суждений экспертов.// Принятие решений и анализ экспертной информации. М., НСКП "Кибернетика", 1988, с.24-44.
4. Нискина Н.Н., Тейман А.И., Шмерлинг Д.С. Алгоритмы непараметрических методов статистики, основанных на рангах. Пакет прикладных программ. Препринт. М., ВНИИСИ, 1986.
5. Чеботарев П.Ю. Метод оценивания объектов по неполному набору парных сравнений.// Комплексный подход к анализу данных в социологии. М., ИСАН СССР, 1988, с.79-93.

СПЕЦИАЛИЗИРОВАННЫЙ ПАКЕТ С Л А М S : КЛАССИФИКАЦИЯ И МНОГОМЕРНОЕ ШКАЛИРОВАНИЕ

Г.А.Сатаров, Е.В.Кулинская, Г.В.Раушенбах

Интерактивный пакет программ прикладного статистического анализа SLAMS представляет собой коммерческую версию I-й очереди Экспертной системы анализа данных (ЭСАД) [1]. Пакет SLAMS включает в себя программы автоматической классификации и многомерного шкалирования, а также средства графического отображения результатов их работы, снабжен базой данных и развитой системой подсказки, что позволяет использовать пакет пользователям с невысокой программистской и статистической подготовкой. В версии 1.00 ППП SLAMS реализованы следующие методы:

- метрическое многомерное шкалирование
- неметрическое многомерное шкалирование по Краскалу
- неметрическое многомерное шкалирование по Джонсону
- метрическое индивидуальное шкалирование с алгоритмом, основанном на разложении Экарта-Янга
- неметрическое индивидуальное шкалирование
- неметрическое многомерное развертывание, модель идеальной точки
- неметрическое многомерное шкалирование, компенсаторно-дистанционная модель
- неметрическое многомерное шкалирование, угловая модель
- метод формальной интерпретации результатов многомерного шкалирования, метод направления максимальной корреляции
- иерархический кластерный анализ
- метод аддитивных деревьев
- метод ISCDATA
- размытый метод ISCDATA
- метод аддитивных кластеров для матриц описания
- метод аддитивных кластеров для матриц близости
- метод Уиндхема построения размытой классификации
- метод построения размытых аддитивных типов
- метод построения размытых параллельных классификаций для бинарных данных

- метод проверки гипотезы об однородности против гипотезы о возможности построения классификации.

Структурными компонентами системы CLAMS являются:

- БАЗА ДАННЫХ (содержит статистические данные, служебные сведения о них и сведения о статистических программах)
- ТАБЛИЧНЫЙ РЕДАКТОР (дает пользователю возможность ввода, просмотра, редактирования данных, описания их структуры и свойств, редукции данных и конструирования на их основе новых данных; осуществляет экспорт-импорт между системой и текстовыми файлами MSDOS)
- ПРОЛОГ (предназначен для запуска статистических программ, автоматического контроля соответствия между программой и данными, ввода управляющих параметров и т.д)
- КОНТРОЛЬНАЯ ТОЧКА (организует взаимодействие пользователя со статистической программой)
- ЭПИЛОГ (организует работу пользователя после завершения работы статистической программы, осуществляет привязку результатов к базе данных, консультирует по дальнейшей обработке полученных данных)
- СПРАВОЧНИК (разъяснение терминов, используемых системой, а также возможностей и смысла действий, доступных пользователю в любой момент диалога с системой).

Пакет CLAMS предназначен для ПЭВМ типа IBM PC/XT, AT с графическим адаптером EGA и работает в операционной среде MS DOS версии 3.3 и выше.

Л и т е р а т у р а

1. Кулинская Е.В., Раушенбах Г.В., Сатаров Г.А. О разработке экспертной системы анализа данных. - В кн.: Проблемы проектирования экспертных систем, ч. II. - М., 1988, с. 160.

О КОМПЛЕКСЕ ПРОГРАММ ДИАНА - ДИСКРИМИНАНТНОГО АНАЛИЗА ВЕКТОРОВ ВЫСОКОЙ РАЗМЕРНОСТИ

В.И.Сердобольский, В.С.Степанов

Комплекс программ ДИАНА разработан в МГУ при участии МИЭМ и филиала ВЭИ. Он предназначен для эффективного решения задачи линейного дискриминантного анализа в условиях большой размерности наблюдений $x \in \mathbb{R}^P$, сравнимой с объемами обучающих выборок $N_\nu : 1 \leq \nu \leq r \sim N_\nu$ ($\nu=1,2$). Комплекс ДИАНА целесообразно использовать в составе одного из статистических пакетов типа SSP, ВМДР, ИПСА и др. [1], хотя он может работать и независимо от них.

В настоящее время ДИАНА реализован на IBM PC/XT в операционной среде MS-DOS на языке ФОРТРАН-77 (версии 3.3 и выше). Основные программы из ДИАНА ранее успешно работали на БЭСМ-6 и ЕС ЭВМ на ФОРТРАН-ГДР, ФОРТРАН-77 и ФОРТРАН-ЕС. Ожидается хорошая работа ДИАНА также и на мини и микроЭВМ.

В основе алгоритмов дискриминантного анализа из ДИАНА лежит линейная дискриминантная функция с параметром z вида

$$\omega_z(x) = \left(x - \frac{\bar{x}^1 + \bar{x}^2}{2} \right)^T S_z^{-1} (\bar{x}^1 - \bar{x}^2), \quad (I)$$

где \bar{x}^ν, S - классические оценки векторов средних μ^ν и ковариационной матрицы Σ популяций \mathcal{N}_ν , $S_z^{-1} = f(S; z)$ - некоторая оценка матрицы Σ^{-1} , построенная с помощью S , x - независимое наблюдение из \mathcal{N}_ν ($\nu=1,2$). Функция (I) получена из классической дискриминантной функции Фишера-Андерсона после замены обычной оценки S^{-1} на регуляризованную S_z^{-1} .

В ДИАНА используются следующие оценки :

$$\text{-ридж-оценка} \quad S_t^{-1} = (I + tS)^{-1}, \quad (2)$$

где I - единичная матрица, $t > 0$ -ридж-параметр ;

-обобщенная ридж-оценка

$$S_\alpha^{-1} = \int_{t \geq 0} S_t^{-1} da(t), \quad (3)$$

где $a = a(t)$ - некоторая функция ограниченной вариации ;

-оценка на основе доминирующих главных компонент

$$S_h^{-1} = S^{-1} U (S - h I) = H \cdot \text{diag} (\tilde{\lambda}_j^{-1} U(\lambda_j - h)) \cdot H^T, \quad (4)$$

где H - ортогональная матрица, $U(t) = \begin{cases} 1, & t > 0 \\ 0, & t \leq 0 \end{cases}$, $\lambda_1, \dots, \lambda_p$ - спектр выборочной матрицы ковариаций S , $h > 0$ - параметр.

Кроме алгоритмов с (2) + (4) для $\nu \geq 2$ классов реализован также ряд классических алгоритмов (ФИШЕР, ЕВКЛИД и др.).

Новизной алгоритмов из ДИАНА является следующее:

1). Реализованы алгоритмы с (2) + (4) и подбором параметра z (t, a, h) из условия

$$z^* = \arg \inf_{z, t} \left(\lim_{p \rightarrow \infty, N_1 \rightarrow \infty} P_{p, N_1}(z) \right), \quad (5)$$

где $P_{p, N_1}(z) = P_{p, N_1}(z / \bar{x}^1, \bar{x}^2, S)$ - условная вероятность ошибок классификации решающего правила $\omega_z(x) \geq c$ с помощью (1).

2). При решении задачи (5) использованы новейшие достижения спектральной теории случайных матриц, поэтому ожидается устойчивость решения (5) к отклонению распределений от $N(\mu, \Sigma)$.

3). Алгоритмы с оценками (2) + (4) с учетом (5) эффективно работают даже при $p \geq (N_1 + N_2)$, а также в условиях мультиколлинеарности системы признаков.

4). Для случая, когда $N_1 \neq N_2$ учитывается дополнительное смещение порога классификации в условиях $p \sim \min(N_1, N_2)$.

5). Для решения задач классификации большой размерности выбран один из самых эффективных алгоритмов решения полной проблемы собственных значений - метод Хаусхолдера с применением QL-алгоритма со сдвигом [2].

Комплекс программ ДИАНА разработан с использованием технологии структурного программирования, имеет модульную структуру. Он состоит из головной программы и 18 подпрограмм и занимает около 2000 строк текста (в исходных модулях).

В заключении сообщения приводится информация об опыте применения комплекса ДИАНА для решения практических задач из области электротехники, медицины и геологии.

Л и т е р а т у р а

1. Сильвестров Д.С. Программное обеспечение прикладной статистики. - М.: Финансы и статистика, 1988.
2. Уилкинсон Д.Р. Справочник алгоритмов на языке Алгол. Линейная алгебра. - М.: Машиностроение, 1976.

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ МЕТОДОВ ГЛАВНЫХ КОМПОНЕНТ И ДИСКРИМИНАНТНОГО АНАЛИЗА В ЗАДАЧАХ БОЛЬШОЙ РАЗМЕРНОСТИ

Н.Н.Товмаченко, О.В.Майстренко, О.С.Малюга

Пусть имеется n независимых наблюдений $\vec{X}_1, \vec{X}_2, \dots, \vec{X}_n$ над случайным вектором $\vec{\xi}$ высокой размерности

$$m_n, M\vec{\xi} = \vec{a}, M(\vec{\xi} - \vec{a})(\vec{\xi} - \vec{a})' = R$$

Для поиска состоятельных оценок конечного числа собственных чисел эмпирической ковариационной матрицы $\lambda_1(\hat{R}), \dots, \lambda_{m_n}(\hat{R})$ не зависящего от n , в случае, если m велико, применяют метод G -анализа [1] G -оценка величины λ_1 равна измеримому решению θ -уравнения

$$\lambda_1(\hat{R}) = \theta(1 - \gamma_n) + \gamma_n \quad \theta^* R_0 [I(\theta + i\varepsilon)(\theta + i\varepsilon)^{-1} - \gamma_n] + (\theta + i\varepsilon) \gamma_n = \frac{m_n}{n}$$

где γ_n -аналитическая функция, являющаяся решением уравнения

$$1 - \gamma_n + \gamma_n \frac{1}{m} \text{Sp} (I_m - \hat{R})^{-1} = (\theta + i\varepsilon)^{-1} \gamma_n$$

$$\hat{R} = n^{-1} \sum_{i=1}^{m_n} (\vec{x}_i - \vec{a})(\vec{x}_i - \vec{a})', \quad \vec{a} = (n+1)^{-1} \sum_{i=1}^{m_n} \vec{x}_i,$$

которое решается методом последовательных приближений.

Пусть $\vec{x}_1, \dots, \vec{x}_n, \vec{y}_1, \dots, \vec{y}_n$, -независимые наблюдения соответственно над случайными векторами $\vec{\xi}$ и $\vec{\eta}$ размерности m , распределёнными по нормальным законам $N(\vec{a}_1, R), N(\vec{a}_2, R)$.

Оценка $(\vec{a}_1 - \vec{a}_2)'(I + R)(\vec{a}_1 - \vec{a}_2)$, где \vec{a}_1, \vec{a}_2, R эмпирические ожидания и ковариационная матрица, не пригодная при решении многомерных задач классификации, так как при увеличении числа компонент $\vec{\xi}$ и $\vec{\eta}$ вырастает число наблюдений n , необходимых для достижения заданной точности оценивания расстояния Махаланобиса.

При некоторых условиях существует асимптотическая нормальная G -оценка регуляризованного расстояния Махаланобиса при условии, что $\lim_{n \rightarrow \infty} \frac{m_n}{n} = \Pi$.

Литература

1. Гирко В.Л. Многомерный статистический анализ.

Киев, Вища шк. 1988

ПАКЕТЫ КЛАССИСТАТ И КЛАССИСТАТ - Д ДЛЯ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ

М.Хоолма, Э.Лелумес, Э.Вийкман, И.Петерсен

КЛАССИСТАТ - это пакет для статистической обработки данных со специально развитыми возможностями анализа на основе группирования по уровням качественных признаков. Пакет предназначен к анализу данных в медицине, социологии, экономике и в других областях. Пакет имеет самостоятельные средства ввода, редактирования и хранения массивов данных. Данные для пакета вводятся в виде прямоугольной таблицы типа объект х признак. В пакете реализована проверка 24 различных по структуре статистических гипотез о независимости 2-5 качественных признаков, анализ соответствующих остатков, проверка гипотез о равенстве частот и об однородности дисперсий, вычисление матриц парных коэффициентов сопряженности между качественными признаками, вычисление разных статистик количественных признаков, таких как среднее, стандартное отклонение, дисперсия, коэффициенты корреляции и регрессии вместе с их доверительными пределами и т.д.

Диалоговая версия пакета, КЛАССИСТАТ-Д, позволяет в течение одного сеанса решать любое количество задач с разными программами пакета. В ходе диалога вводятся данные и параметры программ с возможностью их многократного пользования. Результаты можно осмотреть на экране дисплея перелистованием кадров вперед и назад.

Пакеты работают на ЭВМ ЕС под управлением операционной системы ОС ЕС 6.1. Диалог ведется с дисплеем ЕС 7927-01.

ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ РАБОТЫ КЛАССИФИКАТОРОВ ПРИ ПРИМЕНЕНИИ НОРМАЛИЗУЮЩИХ ТРАНСФОРМАЦИЙ

К.Чепонис

На 19 реальных данных был проведен ряд экспериментов. Для классификации и сравнения использовались алгоритмы классификации из пакета СОРРА-2: классификатор, использующий непараметрические парзеновские оценки (Р), классификатор евклидова расстояния (Е), линейная дискриминантная функция Фишера (F), квадратичная дискриминантная функция (Q). В последующих экспериментах, перед использованием параметрических классификаторов Е, F, Q, были применены четыре трансформации, приближающие распределение исходных признаков к нормальному распределению: обратные "нормализующие" трансформации Джонсона, три трансформации из семейства Бокса-Кокса-Тьюки: $\ln X$, \sqrt{X} , $\sqrt[3]{X}$, являющиеся "нормализующими" для логнормального, пуассоновского и χ^2 -распределений [1]. Оценивание вероятности ошибочной классификации (ВОК) производилось методом скользящего экзаме́на. В табл. I даны результаты экспериментального сравнения (ВОК дана в процентах, СКО – среднеквадратическое отклонение в процентах, ООВ – объём обучающей выборки, Р – число признаков, К – число классов, № – номер данных). Эксперименты показали, что неправильный выбор трансформации (для конкретного классификатора) может увеличить ВОК в 2–3 раза, а правильный выбор – уменьшить ВОК 2–3 раза, однако в наших экспериментах это наблюдалось довольно редко.

Л и т е р а т у р а

1. Чепонис К.А. Обзор нормализующих трансформаций одномерной случайной величины, используемых в анализе данных. – В сб.: Стат. пробл. управления, вып. 74, Бильнюс, 1986.
2. Чепонис К.А. Экспериментальное сравнение трансформаций, приближающих распределения исходных признаков к нормальному распределению в дискриминантном анализе, используя реальные данные. – В сб.: Стат.пробл.упр., вып.74, В., 1986.

Таблица I

Результаты экспериментального сравнения трансформаций и алгоритмов классификации

№	K	P	008	X					√X			³√X			ln X			lg(X)		
				E	F	G	P	CKO	E	F	G	E	F	G	E	F	G	E	F	G
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1.	2	5	200	39.0	41.5	47.5	43.0	3.5	39.0	40.0	45.5	38.0	39.5	44.0	36.5	39.0	42.5	37.0	38.5	41.5
2.	2	4	200	44.0	35.5	42.5	43.5	3.5	44.5	34.5	41.5	41.0	33.5	40.5	36.0	33.0	39.0	40.5	41.5	40.5
3.	2	5	45	20.0	15.6	13.3	15.6	5.4	24.4	13.3	11.1	37.8	13.3	11.1	20.0	15.6	11.1	15.6	20.0	17.8
4.	2	5	200	19.5	19.0	21.5	16.0	2.6	15.0	20.0	20.0	19.5	19.0	21.5	16.5	24.5	17.5	-	-	-
5.	2	5	242	13.6	16.1	15.3	3.7	1.2	22.3	17.4	16.5	34.3	17.4	25.6	15.3	17.8	27.7	14.5	11.7	18.6
6.	2	5	200	13.5	9.0	9.0	0.0	2.0	9.0	9.0	9.0	9.0	9.0	9.0	9.0	9.0	9.0	14.0	9.0	4.5
7.	3	5	150	46.7	53.3	49.3	38.0	4.0	46.0	51.3	50.7	46.0	50.0	52.7	49.3	46.0	54.7	42.7	44.0	57.3
8.	2	5	300	10.3	7.0	10.3	6.7	1.4	8.3	8.0	7.0	7.3	8.3	6.7	10.3	7.7	7.3	7.0	6.3	6.3
9.	4	5	400	71.8	68.3	69.5	69.8	2.3	72.0	68.8	70.5	73.8	69.3	70.0	73.0	68.8	71.0	66.0	70.5	69.3
10.	5	4	555	58.7	30.1	29.4	18.2	1.6	50.6	29.7	26.9	47.9	30.1	27.8	44.7	27.0	27.6	50.5	29.6	26.7
11.	3	5	247	59.5	30.4	3.6	1.2	0.7	49.8	25.5	5.7	48.2	25.5	6.1	51.4	27.1	7.7	59.1	29.6	17.0
12.	2	5	179	16.8	13.4	23.5	4.5	1.5	15.6	15.1	20.1	15.1	15.6	20.1	16.2	14.5	20.1	27.4	14.5	11.7
13.	4	5	280	31.1	8.9	9.6	8.6	1.7	28.9	17.1	8.0	29.3	8.2	8.9	29.6	11.8	8.2	28.9	16.1	10.0
14.	2	5	117	41.9	6.0	13.4	6.0	1.6	43.8	6.0	13.4	10.6	5.5	13.4	11.5	5.5	13.4	11.5	6.5	11.5
15.	2	5	200	47.5	33.0	39.5	36.5	3.4	49.5	37.0	35.5	48.0	37.0	34.5	46.0	33.5	32.5	38.5	38.0	46.0
16.	2	5	122	37.7	32.8	34.4	30.3	4.2	29.5	36.1	38.5	31.2	38.5	39.3	33.6	38.5	40.1	27.9	32.8	30.3
17.	3	5	300	37.7	32.3	26.0	20.0	2.3	39.3	33.3	26.3	42.3	32.3	26.7	47.3	33.7	30.0	33.7	34.3	31.0
18.	2	1	23	21.7	26.1	34.8	21.7	8.6	21.7	26.1	34.8	21.7	26.1	34.8	21.7	26.1	34.8	21.7	26.1	26.1
19.	6	3	180	51.7	50.6	47.2	44.4	3.7	48.9	47.8	48.3	47.2	47.2	48.9	51.1	48.3	48.3	58.3	50.0	47.2

ПРОБЛЕМНО-ОРИЕНТИРОВАННЫЕ ПРОЦЕДУРЫ СТАТИСТИЧЕСКОГО
АНАЛИЗА ПОКАЗАТЕЛЕЙ СРЕДСТВ ИНФОРМАТИКИ

В.О.Гречко, И.В.Жовтяк, А.М.Захарин

Диалоговая проблемно-ориентированная система обработки данных "Уровень" (ДС "Уровень") представляет собой комплекс программных средств, предназначенных для автоматизации сбора и обработки информации о технико-экономических показателях средств информатики (СИ). Ее программные средства реализованы на языке "Си" и функционируют в среде MS/DOS ПЭВМ типа IBM PC. ДС "Уровень" является составной частью, создаваемой в ИК АН УССР, автоматизированной системы маркетинга СИ. ДС "Уровень" позволяет осуществлять анализ информации на всех стадиях жизненного цикла изделия с целью выработки рекомендаций для принятия управленческих решений. Система реализует комплексный подход к оценке технико-экономического уровня (ТЭУ) СИ по показателям, отражающим техническое совершенство, качество и экономические характеристики изделия. Основными компонентами ДС "Уровень" являются: инструментальные средства управления данными и организации диалога, пакет прикладных программ статистической обработки, совокупность диалоговых сценариев обслуживания пользователей.

Инструментальные средства, используемые в ДС "Уровень", обеспечивают ведение информационной базы сетевого типа, включающей набор справочных функций, реализующих ограниченные возможности реляционной алгебры на связях. Диалоговый инструментарий содержит операции форматированного вывода на экран и получения отчетов пользователей с предварительным контролем данных, а также различные формы организации "меню". Обращение пользователей к прикладным программам осуществляется через совокупность диалоговых сценариев, которые обеспечивают выбор объектов и средств обработки, предназначенных для решения задач статистического анализа характеристик данного класса СИ; формирования схем счета для вычисления ТЭУ (включая верификацию и оценку погрешности); классификации и выбора аналога; прогнозирования технико-экономических характеристик.

Статистический анализ включает исследование парных, частных и множественных коэффициентов корреляции, агрегирование информации путем выделения общих факторов и наиболее информативных показателей, разбиение всей совокупности показателей на непересекающиеся классы коррелированных и классы некоррелированных показателей, которые могут быть использованы как предикторные при построении уравнений регрессии. Результаты, полученные при решении данной задачи, используются при прогнозировании и оценке ТЭУ.

ТЭУ СИ предлагается рассматривать как характеристику его свойств, определяемых интересами потребителя. Расчетные методы оценки разделяют на статистические, апромасимационные и методы распознавания образов. С нашей точки зрения наиболее целесообразным является применение методов первого типа, так как: многие показатели имеют случайный характер и можно говорить лишь о статистических оценках значений. Кроме того, при таком подходе появляются дополнительные возможности учета структуры исходных данных и оценки погрешности получаемых величин. В зависимости от результатов статистического анализа исходных данных в системе применяются различные схемы счета ТЭУ.

Одной из особенностей системы является использование для решения вышеуказанных задач общих статистических процедур, что позволяет расширить область применения системы без существенного увеличения ее размеров. Отличия при их использовании возникают на этапах подготовки репрезентативной выборки и интерпретации полученных результатов. Так, например, при оценке ТЭУ формирование совокупностей изучаемых объектов (базы сравнения) осуществляется в соответствии с системой ограничений, задающих подмножество допустимых значений параметров. В прогнозировании же используются динамических ряды значений показателей.

В перспективе предполагается дальнейшее развитие функциональных возможностей системы, с целью использования ее для анализа и прогнозирования конкурентноспособности СИ, что приводит к необходимости более детального исследования рынка СИ.

БАЙЕСОВСКИЙ ПОДХОД К СТАТИСТИЧЕСКОМУ КОНТРОЛЮ КАЧЕСТВА ПРИ ИСПОЛЬЗОВАНИИ ПРИНЦИПА РАСПРЕДЕЛЕНИЯ ПРИОРИТЕТОВ

В.А.Лапидус

При байесовском подходе к статистическому приемочному контролю качества выбор планов осуществляется на основе минимизации средней безусловной функции потерь, учитывающей потери от ошибочных решений и затраты на контроль. При этом используются априорные распределения случайных величин — показателей качества отдельных партий, что позволяет повысить эффективность контроля, в том числе уменьшить объем необходимых экспериментов. На практике применение байесовской схемы встречается определенные трудности, обусловленные правовыми аспектами. А именно: поставщик и потребитель (инспектор госприемки), элементом производственных отношений которых является приемочный контроль качества, в силу несовпадающих интересов по разному относятся к априорной информации и по разному оценивают ее достоверность. Степень их информированности тоже различна: поставщик лучше знает (во всяком случае может и обязан лучше знать) качество изготавливаемой им продукции, но потребитель чаще всего отказывается признавать информацию, сообщаемую поставщиком.

Это приводит к тому, что байесовские методы, несмотря на свои теоретические преимущества, фактически не применяются на практике.

Изменить положение в области применения байесовских методов может подход к статистическому контролю на основе принципа распределения приоритетов (ПП) [1]. Устанавливая четкие правовые приоритеты, он тем самым позволяет индивидуально учесть имеющуюся у сторон информацию о качестве продукции.

В соответствии с ПП при контроле, осуществляемом по-

ставщиком, последний имеет право в одностороннем порядке выбирать планы контроля из числа удовлетворяющих ограничениям на условный риск потребителя β [2].

В связи с этим поставщик при выборе планов может использовать всю имеющуюся у него априорную информацию без согласования ее с потребителем.

Потребитель учитывает свою априорную информацию при назначении риска β .

Выбор β осуществляется из формулы

$$\beta = \begin{cases} \frac{\beta\delta}{\rho}, & \text{если } \frac{\beta\delta}{\rho} < \beta_m, \\ \beta_m, & \text{если } \frac{\beta\delta}{\rho} \geq \beta_m, \end{cases}$$

где $\beta\delta$ – байесовский риск потребителя, устанавливаемый в стандартах;

ρ – априорная оценка вероятности не выполнения требований к качеству;

β_m – предельное (регламентируемое) значение условного риска потребителя.

Учет априорной информации потребителя в виде значения ρ (при малых ρ) позволяет существенно уменьшить объемы контроля по сравнению с параметрическим подходом.

В работе рассмотрена байесовская постановка задачи выбора оптимальных планов контроля, показано, в каких случаях она сводится к параметрической, сообщается о программном и методическом обеспечении.

Литература

1. Липидус В.А. Контроль качества на основе принципа распределения приоритетов. НКК-1984, №6.
2. Липидус В.А., Розно М.И., Фүфаев А.А. и др. Стандартизация методов контроля качества и надежности в машиностроении: Обзорная информ. (Сер. Стандартизация); Вып. I ВНИИКИ, М.: 1987. – 68 с.

КОМПЛЕКС ПРИКЛАДНЫХ ПРОГРАММ
ПО АНАЛИЗУ ЦЕНЗУРИРОВАННЫХ ДАННЫХ (POD)

А.П.Макаров, Ю.К.Беляев,
В.А.Беляева, Т.Н.Дугина

Предлагаемый комплекс прикладных программ предназначен для анализа данных типа времени жизни (Survival analysis), [1]. Это, как правило, цензурированные данные, которые возникают при испытаниях изделий на надежность или в процессе их эксплуатации, в медико-биологических исследованиях и во многих других ситуациях, [2, 3].

В основу комплекса программ POD положены проводимые авторами исследования по изучению моделей накопления неполных данных, [2-7]. Заметим, что во многих приложениях процесс накопления данных может быть формализован с помощью порций, [2,3]. Порция $X_j = (S_{ij}, \Gamma_{ij}, i=1, \dots, I_j)$ - есть набор случайного количества I_j , вообще говоря, зависимых неотрицательных случайных величин S_{ij} , снабженных метками Γ_{ij} . Случайные величины S_{ij} являются наблюдаемыми наработками (временами жизни), а соответствующие им метки Γ_{ij} кодируют причину прекращения наблюдения (отказ).

Комплекс программ POD состоит из 4 разделов. Первый раздел составляют программы, которые из порций данных образуют специального вида порядковые таблицы, играющие роль интерфейса для основных программ. Второй раздел составляют базовые программы расчета множительных оценок, доверительных оценок и полос для функций распределения; ядерных оценок для функций интенсивности отказов с выбором оптимальной (по критерию минимального среднеквадратического отклонения) ширины окна и др. В третий раздел включены вспомогательные программы, в их числе программы, обеспечивающие выдачу (на

широкую печать, экран) таблиц и графиков. В этот же раздел включены некоторые типовые программы с целью обеспечения независимости комплекса программ POD от других пакетов. Заключительный раздел моделирования наряду с выработкой данных, соответствующих различным моделям, дает возможности использования методов эмпирического моделирования (бутстреп-па).

Завершенная версия комплекса программ POD написана на языке FORTRAN и ориентирована на ЭВМ типа ЕС (например, ЕС-1045). В настоящее время продолжается работа по расширению возможностей комплекса программ POD и его адаптации на персональные компьютеры.

Л и т е р а т у р а

1. Кокс Д.Р., Оукс Д. Анализ данных типа времени жизни. М., Финансы и статистика, 1988.
2. Беляев Ю.К. Статистические методы обработки неполных данных. М., Знание, 1987.
3. Беляев Ю.К. Множительные оценки вероятности безотказной работы. Изв.АН СССР, Техн.кибернетика, 1985, № 4, с.45-49.
4. Макаров А.П. Асимптотические свойства апостериорных распределений в задачах накопления неполных данных. Автореферат диссертации на соискание ученой степени к.ф.м.н., М., 1987.
5. Чистякова Н.В. Асимптотические свойства непараметрических оценок распределений длин требований в некоторых системах массового обслуживания. Автореферат диссертации на соискание ученой степени к.ф.м.н. М., 1988.
6. Макаров А.П., Беляев Ю.К., Беляева В.А., Дугина Т.Н. Математическое обеспечение обработки цензурированных данных. Тезисы III Всесоюзной научно-технической конференции, г. Тарту, 18-19 сентября 1985, II, с.184-186.
7. Беляев Ю.К., Макаров А.П. Статистический анализ массивов неполных данных. Тезисы докладов III Всесоюзной школы-семинара, Цахкадзор, окт. 1987, с.185-186.

ПРИМЕНЕНИЕ СТАТИСТИЧЕСКОГО АНАЛИЗА В КЛИНИЧЕСКОЙ ИММУНОЛОГИЧЕСКОЙ ЛАБОРАТОРИИ

Е.Н.Наумова, Т.В.Елисеева, В.В.Губарев, М.П.Цапенко,
И.Б.Белан

Для повышения эффективности применения статистического анализа клинических и лабораторных данных в иммунологии, решения проблемы выбора методов, адекватных специфике эмпирического материала, разработан аппарат формализованного описания технологических процессов иммунологических экспериментов.

Иммунологический эксперимент, в ходе которого выполняются манипуляции над биологическим объектом и результатом которого являются показатели функционирования иммунной системы, представлен в виде технологического процесса и описан с помощью содержательных логических схем алгоритмов /1/, дополненных специализированными операторами, отражающими сущность технологической операции (ТО). На основе анализа реальных экспериментальных схем каждой ТО поставлено в соответствие некоторое преобразование случайной величины (аддитивное или мультипликативное засорение, линейное или нелинейное преобразования, усечение и т.п.). Такое формализованное описание позволило провести анализ вероятностных свойств выборочных данных исходя из физики реального процесса. Осуществлена программная реализация формализованного описания и анализа погрешностей в виде комплекса программ IMLSA.

Построение вероятностных моделей одномерных унимодальных распределений результатов иммунологических измерений выполняется с помощью комплекса программ IMSTAT, в котором на основе данных предыдущего этапа работы программы IMLSA осуществляется выбор методов статистического анализа, в

частности, в разделе анализа взаимосвязей, предлагается использовать векторные характеристики связи (наборы из известных числовых характеристик корреляционного типа) /2/.

С помощью комплексов программ IMISA и IMSTAT проводится контроль качества проведения иммунологического эксперимента и осуществляется проверка степени овладения методикой обучающегося специалиста, а также обучение иммунологов курсу математической статистики.

Программное обеспечение выполнено в виде автоматизированных рабочих мест специалистов-иммунологов для персональных компьютеров типа ДВК на языках БЭЙСИК и ФОРТРАН, ориентировано на неподготовленного пользователя, имеет широкие сервисные возможности по организации и преобразованию создаваемых в ходе экспериментальных и клинических работ локальных баз данных, успешно работает в условиях клинической иммунологической лаборатории.

Помимо этого осуществлена аппаратно-программная межмашинная связь ДВК-2М с "Электроникой-60", управляющей микроскопом-фотометром "Люмам-ПМ II" и имеющей в своем составе перфоленточное устройство ввода-вывода. Выполненная разработка реализует возможность дополнительного использования ввода-вывода на перфоленту для ДВК, не влияя на автоматизированное проведение измерений на микроскопе. Причем комплекс программ IMSTAT позволяет иммунологу с микроскопа автоматически переносить результаты измерений в ДВК для их статистической обработки, сбора, преобразования и хранения данных на гибких магнитных дисках.

Л и т е р а т у р а

1. Цапенко М.П. Измерительные информационные системы. М., Энергоатомиздат, 1985.
2. Наумова Е.Н. Разработка математического обеспечения вероятностного описания показателей функционирования иммунной системы / Автореф. канд. дис., Новосибирск, 1988.

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ARFA
(АВТОМАТИЗИРОВАННЫЙ РЕНТГЕНОФАЗОВЫЙ АНАЛИЗ)

В.Пивоврюнас

Рентгеновский фазовый анализ (РФА) является одним из основных методов исследования состава поликристаллических материалов и широко используется для оценки качества продукции. Сущность РФА заключается в следующем. В результате проведения исследований на дифрактометре регистрируется дифрактограмма (рентгеновский спектр). В общем случае – это функция $y(x)$ определяющая величину интенсивности y в точке x . Каждая кристаллическая фаза (вещество) дает всегда одинаковый рентгеновский спектр. Множество таких эталонных спектров фаз хранится в справочниках и картотеках. Рентгеновский спектр материала является суперпозицией эталонных спектров, составляющих этот материал фаз. Таким образом спектр смеси можно описать следующим образом:

$$y(x) = \sum_{i=1}^M c^{(i)} y_0^{(i)}(x), \quad (I)$$

где $y_0^{(i)}(x)$ – функция, характеризующая эталонный спектр; $c^{(i)}$ – коэффициент, пропорциональный концентрации эталона в материале (условная концентрация); M – число фаз; i – индекс эталонного спектра.

Качественный РФА, т.е. идентификация веществ, составляющих исследуемый материал, осуществляется путем сопоставления спектра анализируемого материала с эталонными спектрами веществ, и отыскивания такого набора последних, спектр которого наилучшим образом соответствует исследуемому объекту. Известно ряд критериев (решающих функций) для оценки степени соответствия между анализируемым и эталонными спектрами. В [1,2] показано, что наиболее распространенные критерии получены аппроксимируя спектральные линии прямоугольниками или равнобедренными треугольниками.

В работе [3] нами предложена и исследована статистическая модель, сущность которой заключается в следующем. Спектр, т.е. положения (вектор X) и интенсивности (вектор Y) линий характеризуется вектором $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$ с известным распределением

вероятностей. Это распределение предполагается нормальным

$Z \sim N(Z_0, \Sigma)$, где $Z_0 = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ - вектор средних, Σ - ковариационная матрица. Здесь x_0 - вектор положений линий эталонных спектров, а компоненты вектора y_0 определяются соответственным уравнению (1).

Для этой модели, с использованием разных предположений о виде ковариационной матрицы Σ , нами разработана группа статистических методов, которые используются для идентификации спектров.

На базе полученных теоретических результатов создан пакет прикладных программ (ППП) ARFA. Пакет разработан на мини-ЭВМ СМ-1600 (объем оперативной памяти 256 КБ, операционная система ОСРВ); программные модули написаны на алгоритмическом языке ФОРТРАН-77.

Машинная идентификация спектров обычно связана с многократным числом попыток анализа. Поэтому основные требования, которые ставились при разработке пакета - это удобность обращения и скорость анализа. ППП оснащен удобным для пользователя входным языком, позволяющим выбрать режим работы, вводить исходные данные и параметры анализа. Создан быстродействующий иерархический алгоритм. Разработано программное обеспечение для эксплуатации базы данных эталонных и анализируемых спектров.

Л и т е р а т у р а

1. Васильев Е.К., Нахмансон М.С. Качественный рентгенофазовый анализ. - Новосибирск, Наука, 1936.
2. Пиворюнас В., Раудис Ш. Обзор методов автоматизированного анализа состава материалов по их рентгеновским спектрам. - В сб.: Статистические проблемы управления. - Вильнюс, 1988.
3. Пиворюнас В. Линейная дискриминантная функция для идентификации спектров. В сб.: Статистические проблемы управления. - Вильнюс, 1978.

МАТЕМАТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА ЭКСПЕРТНОЙ И ЭКОНОМИЧЕСКОЙ ИНФОРМАЦИИ ПРИ ФОРМИРОВАНИИ ГОСУДАРСТВЕННОГО ЗАКАЗА

Г.В.Раушенбах, В.Д.Воскобойкин, А.А.Заславский, В.Б.Митрофанов

В настоящее время в нашей стране приобретают распространение новые организационные формы управления наукой, опирающиеся на использование конкурсного механизма для отбора наиболее перспективных исследований и разработок (ИР) [1]. При подготовке и проведении подобных конкурсов возникает ряд нетривиальных задач обработки многокритериальной экспертной информации о проектах ИР, анализа согласованности мнений экспертов, поддержки решений с надежностью оценки приоритета отдельных ИР и об упорядочении их совокупности. Решение этих задач должно быть соотнесено с анализом доступной экономической информации и послужить основой для распределения материальных ресурсов государственного заказа по развитию науки и техники (ГЗ), научно-технической программы (НТП) и т.п. Апостериорный анализ экспертной информации должен также дать возможность уточнить области максимальной компетенции членов экспертных советов (ЭС) и комиссий экспертов (ЭК) с целью наиболее эффективного использования их знаний при контроле за выполнением ИР и проведении последующих конкурсов.

Для автоматизации процессов сбора обработки и хранения информации при формировании ГЗ (НТП) нами разработана система информационно-вычислительного сопровождения конкурсов ГРАНТ, реализованная на IBM-подобных персональных ЭВМ. Коммерческие версии системы ГРАНТ включают до 5 взаимосвязанных подсистем, к числу которых относятся подсистемы поддержки принятия решений "Конкурс", "Ресурс - 1", "Ресурс - 2", интеллектуальная ИПС "эксперт" и экспертная система анализа данных [2] (версия 1.1 - ИСПП СЛАНС). Для большинства из названных подсистем предусмотрена возможность их автономного использования,

что позволяет формировать структуру системы ГРАНТ в соответствии с положениями заказчика. При анализе экспертной информации в системе ГРАНТ используются как оригинальные, так и традиционные методы анализа нечисловых данных, включающие в качестве обязательного условия статистический анализ надежности итоговых решений ЭС / КЭ. Для углубленного анализа данных используются методы кластер - анализа и многомерного шкалирования, реализованные в **III CLAMS**.

Система ГРАНТ и входящие в ее состав подсистемы внедрены в ряде ЭС различных ведомств и межведомственных организаций и используются при формировании ГЗ в промышленных и непромышленных отраслях народного хозяйства, при формировании государственных и региональных научно - технических программ и т.д. В докладе приводятся результаты анализа материалов конкурсов, состоявшихся в 1988 - 1989 г.г. Исследования проводились при поддержке советско- американского фонда "Культурная инициатива".

Л и т е р а т у р а

1. Раушенбах Г.В., Ханыков А.В. Экспертные методы оценки и отбора приоритетных исследований и разработок в биотехнологии. - Биотехнология, 1988, т.4, с.668 - 576.
2. Кулинская Е.В., Раушенбах Г.В., Сатаров Г.А.. О разработке экспертной системы анализа данных. - В кн.: Проблемы проектирования экспертных систем, ч. II. - М., 1988, С.160

ОБЗОР ПОПУЛЯЦИИ ОТЕЧЕСТВЕННОГО СТАТИСТИЧЕСКОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Ю.А.Великий

В докладе представлены обзорные данные по отечественному статистическому программному обеспечению (СПО). Данные получены в результате проведения первого этапа экспертного обследования СПО, проводимого СКБ "Спектр" Киевского госуниверситета и начатого в [1].

Описаны таксономические группы популяции выявленных статистических программных средств по системной организации (статистические комплексы /КП/ и библиотеки программ; статистические пакеты прикладных программ /ППП/; статистические макросистемы и др.), проблемной ориентации (системы управления базами данных, ориентированные на задачи статистического анализа; специализированные ППП и КП для редактирования и табулирования данных; ППП для общего статистического анализа данных; ППП и КП для статистического анализа данных выборочных обследований и др.), типам ЭВМ и другим признакам.

Л И Т Е Р А Т У Р А

- I. Сильвестров Д.С. Программное обеспечение прикладной статистики: Обзор состояния. Тенденции развития. - М.: Финансы и статистика, 1988. - 240 с.

ТЕСТИРОВАНИЕ ПРОЦЕДУР ДИСПЕРСИОННОГО АНАЛИЗА

МЕТОДОМ ТОЧНЫХ ВЫБОРОК

М. Вяхи

Удобным средством тестирования статистических процедур является метод точных выборок (см. [1]). Опишем точную выборку для сбалансированного двухфакторного дисперсионного анализа.

Более подробный обзор конструирования подходящих для тестирования процедур дисперсионного анализа выборок приведен в статье [2].

При двухфакторном дисперсионном анализе используются для конструирования точных выборок следующие произвольно задаваемые параметры: количество измерений в группе (обозначим через n , причем для упрощения конструкции введем требование, что n - нечетное число), количество уровней I фактора (k_1) и количество уровней II фактора (k_2), параметр внутригрупповой расхожимости (u), параметры определяющие средние значения групп (d_1 и d_2), параметр взаимодействия (b).

Группой назовем значения признак-функции, соответствующие определенной комбинации (i, j) уровней факторов. Притом первую группу составляют значения, которые соответствуют комбинации уровней факторов $(1, 1)$ и последнюю группу значения, соответствующие комбинации (k_1, k_2) .

При конструировании точной выборки исходим из "основной структуры" группы, которая составляется следующим образом:

$$-\frac{n-1}{2}, -\frac{n-1}{2}+1, \dots, \frac{n-1}{2}.$$

Первому уровню первого фактора соответствующие значения получим умножением элементов "основной структуры" на параметр u . h -тому уровню первого фактора соответствующие данные получим при сложении значениям первого уровня величины $(h-1)d_1$. Полученную выборку будем считать данными, соответствующими первому уровню второго фактора. Данные, соответствующие i -тому уровню второго фактора, получим слагая элементы первого уровня с величиной $(i-1)d_2$. Для получения взаимодействия прибавим к первой и последней группам констант b .

Общий член приведенной конструкции можно представить в

виде формулы:

$$y_{hij} = -\frac{n-1}{2} u + u(j-1) + d_1(i-1) + d_2(h-1) + b(\delta_{h1} \delta_{i1} + \delta_{hk_1} \delta_{ik_2}),$$

где $j=1, \dots, n$; $i=1, \dots, k_1$; $h=1, \dots, k_2$; δ_{ij} - символ Кронекера.

Проверяемые статистики вычисляются следующим образом:

а) средние

$$\bar{y} \dots = \frac{1}{2} (d_1(k_1-1) + d_2(k_2-1)) + \frac{2b}{k_1 k_2},$$

$$\bar{y}_{h \dots} = \frac{d_1(k_1-1)}{2} + (h-1)d_2 + \frac{b}{k_1} (\delta_{h1} + \delta_{hk_2}),$$

$$\bar{y}_{\dots i} = d_1(i-1) + \frac{d_2(k_2-1)}{2} + \frac{b}{k_2} (\delta_{i1} + \delta_{ik_1}),$$

$$\bar{y}_{hi \dots} = d_1(i-1) + d_2(h-1) + b(\delta_{h1} \delta_{i1} + \delta_{hk_2} \delta_{ik_1});$$

б) компоненты дисперсии

$$S_1^2 = \frac{d_1^2 N(k_1^2-1)}{12} + \frac{2b^2 n(k_1-2)}{k_1 k_2},$$

$$S_2^2 = \frac{d_2^2 N(k_2^2-1)}{12} + \frac{2b^2 n(k_2-2)}{k_1 k_2},$$

$$S_{1,2}^2 = \frac{2b^2 n^2 ((k_2-1)(k_1-1)+1)}{N},$$

$$S^2 = \frac{u^2 N(n^2-1)}{12};$$

с) F-статистики

$$F_1 = \frac{d_1^2 k_1 k_2 (k_1+1)}{u^2 (n+1)} + \frac{24b^2 (k_1-2)}{u^2 k_1 k_2 (k_2-1)(n+1)},$$

$$F_2 = \frac{d_2^2 k_1 k_2 (k_2+1)}{u^2 (n+1)} + \frac{24b^2 (k_2-2)}{u^2 k_1 k_2 (k_2-1)(n+1)},$$

$$F_{1,2} = \frac{24b^2 n ((k_2-1)(k_1-1)+1)}{u^2 N(k_2-1)(k_1-1)(n+1)}.$$

Л и т е р а т у р а

1. Tiit, E.-M. Exact samples of multivariate distributions and their exploitation in statistical algorithm's testing. Acta et Commentationes Universitatis Tartuensis, 1986, Nr. 733, p. 40-62.

2. Vähi, M. Exact samples for testing ANOVA procedures. Acta et Commentationes Universitatis Tartuensis, (в печати).

СИСТЕМА ЭКСПЕРТНО-СТАТИСТИЧЕСКОЙ ОЦЕНКИ КАЧЕСТВА ПРОГРАММНЫХ СРЕДСТВ (ЭСОК ПС)

В.П.Морозов, И.П.Колпаков

В работе рассматривается методика сравнительного анализа и оценки ПС с использованием методов факторного анализа для выделения агрегированных показателей качества и их интерпретации. Под качеством понимается совокупность функциональных (количественных) и социальных (качественных) показателей, характеризующих процессы разработки/эксплуатации ПС.

Ядром ЭСОК является информационный фонд, содержащий экспертные оценки ПС в соответствии с системой показателей, охватывающей наиболее распространенные классы ПС (информационные системы, СУБД, ППП решения экономических задач и т.п.). Система показателей в основном согласуется с [1].

Оценка социальных показателей (а также функциональных показателей, истинное значение которых установить проблематично) реализуется группой экспертов на основе метода взвешивания балльных оценок [2].

Система агрегированных показателей представляет собой выделенные и интерпретированные факторы, полученные путем применения методов факторного анализа к экспертно-статистической информации, хранящейся в информационном фонде.

Работы по анализу и оценке качества ПС осуществляются в рамках целевой программы по качеству ГКНТ СССР.

Программное обеспечение ЭСОК разрабатывается на базе персональной ЭВМ класса ЕС-1840.

Л и т е р а т у р а

1. РТМ 25 690-85. Программные средства. Оценка качества. -М.: Минприбор, 1985
2. Экспертные методы оценки качества программных средств. -Калинин: Центрпрограммсистем, 1987

ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ АЛГОРИТМОВ ПРИКЛАДНОЙ СТАТИСТИКИ

М.Л.Петрович, М.И.Давидович, Т.И.Левина,
Д.А.Мацевич, А.Н.Шитик, Г.К.Шлег

Создание библиотеки прикладных программ (БИМ-М) по различным разделам прикладной статистики требует отбора численных методов, позволяющих получить надежные результаты на ЭВМ различных типов для исходных данных, измеренных в различных шкалах и имеющих особенности.

В докладе рассматриваются результаты численного исследования на ЭВМ алгоритмов:

- вычисления функций распределения вероятностей и их квантилей [1,2] ;
- анализа структуры и тесноты статистической связи между исследуемыми переменными [3] ;
- оценивания параметров, линейно входящих в уравнение регрессии [4] ;
- проверки статистических гипотез (основанных на нормальном распределении и свободных от вида распределения вероятностей) [3,5] .

Л И Т Е Р А Т У Р А

1. Давидович М.И. Вычисление функций распределения и их квантилей. - Мн.: 1988. - 50 с. - (Препринт / АН БССР. Ин-т математики; № 1 (311)).
2. Петрович М.Л., Давидович М.И. Статистическое оценивание и проверка гипотез на ЭВМ. - М: Финансы и статистика, 1988.
3. Программное обеспечение ЭВМ: Библиотека прикладных программ БИМ-М / АН БССР. Ин-т математики. - Мн.: 1989. Прикладная статистика. Корреляционный анализ. - 150 с.
4. Петрович М.Л., Шлег Г.К. Робастная регрессия: оценки и сравнение методом Монте-Карло // Заводская лаборатория, № 3, 1987. - с. 41-48.
5. Петрович М.Л., Левина Т.И., Шитик А.Н., Шлег Г.К. Программное обеспечение ЭВМ / АН БССР. Ин-т математики. - Мн.: 1988 - Вып. 8. - 231 с.

ИССЛЕДОВАНИЕ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК ЭКСПЕРИМЕНТАЛЬНОГО ВЫБОРА ОПТИМАЛЬНОГО АЛГОРИТМА ОБРАБОТКИ ДАННЫХ

В.Пикялис

Широкое распространение вычислительной техники, создание большого количества алгоритмов обработки данных привело к тому, что сейчас часто одну и ту же задачу можно решать различными методами. Для выбора "наилучшего" метода обычно применяется способ испытания на ЭВМ различных конкурирующих методов. Совокупность результатов такого испытания (оценок качества алгоритмов K_i) можно рассматривать как случайную величину. В [1] отмечалось, что число возможных вариантов даже при умеренном объеме выборок ($m = 20 + 50$) и ограниченном объеме генеральной совокупности результатов испытаний, скажем $N = 200 + 600$, весьма большое. Поэтому предлагался способ оценки среднего значения экстремального качества варианта, основанный на статистическом моделировании.

Однако, хотя число вариантов выборок очень большое, число различных экстремальных значений меньше N . Если рассматривать вариационный ряд $K_1 < K_2 < \dots < K_N$, то при переборе всех возможных выборок объема m частота, с которой K_j появляется в качестве минимального значения, есть $w_j = C_{N-j}^{m-1} / C_N^m$. Это позволяет вычислять характеристики распределения кажущихся, фактических и идеальных значений оценок качества метода обработки данных.

В докладе приводятся и обсуждаются графики зависимости средних значений и дисперсий оценок качества от объема выборки при различных условиях эксперимента.

Л и т е р а т у р а

1. Раудис Ш., Пикялис В., Стасайтис Д. Влияние числа исходных и выбираемых признаков, зависимости между признаками и типа классификатора на точность выбора признаков. Статистические проблемы управления, вып. 74, Вильнюс, 1986.

МЕТОДИКА ЭКСПЕРТНОГО ОБСЛЕДОВАНИЯ СТАТИСТИЧЕСКОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Д.С. Сильвестров

В докладе представлена методика экспертного обследования отечественного статистического программного обеспечения (СПО).

В основу методики положены принципы оценки статистических программ АСА [1] и предложения по ее модификации применительно к отечественной практике и современному этапу развития СПО, изложенные в [2].

Обсуждаются перечень характеристик, шкалы для оценки уровня реализации возможностей статистических программных средств по различным группам характеристик, интегральные критерии оценки качества СПО.

Представлен проект экспертного вопросника.

Описаны процедуры проведения экспертного обследования.

Обсуждается проект создания и сопровождения компьютерной базы данных и экспертной системы оценки качества СПО.

Л и т е р а т у р а

1. Francis I. Statistical Software: A comparative review. - Amsterdam - New York: North-Holland, 1981.
2. Сильвестров Д.С. Программное обеспечение прикладной статистики: Обзор состояния. Тенденции развития. - М.: Финансы и статистика, 1988. - 240с.

ТЕСТИРОВАНИЕ МАТОБЕСПЕЧЕНИЯ ПО СТАТИСТИКЕ - ТЕОРЕТИЧЕСКИЕ И МЕТОДОЛОГИЧЕСКИЕ ПРОБЛЕМЫ

Э.-М. Тийт

С ростом числа программных средств по статистике (ПСС) и повышением степени их автоматизации (например, в экспертных системах), возникает серьезная задача тестирования и оценивания ПСС (см. [2]).

Если многие свойства ПСС (напр. объем области применения, простота пользования и т. д.) можно оценить при помощи экспертов (см. [1]), то свойства, связанные с математической корректностью и точностью ПСС требуют специального обследования.

По существу, каждое ПСС A вычисляет на основании статистического материала (выборки) X значение $\beta(A, X)$ некоторому параметр-вектору β , истинное значение $\beta(X)$ которого, в общем, неизвестно. Множество разностей

$$\{\beta(A, X) - \beta(X) : X \in \mathcal{X}\},$$

где \mathcal{X} множество всевозможных обрабатываемых выборок, и характеризует качество ПСС A .

Для оценивания неизвестного $\beta(X)$ на практике пользуются следующими приемами:

1. Считается некое ПСС A_0 "абсолютно правильным" и выбирается $\beta(X) := \beta(A_0, X)$.

2. Рассматриваются "классические выборки" X_0 (например ирис-данные Фишера, для которых $\beta(X_0)$ многократно вычислены и считаются известными).

3. Генерируются случайные выборки, имеющие заданное распределение P и вычисляется $\beta(X)$ теоретически по распределению P .

В настоящем докладе предлагается новый подход: построение таких комплектов исходных данных X , т. н. точных выборок, для которых значение $\beta(X)$ вычисляется автоматически абсолютно точно.

ЛИТЕРАТУРА

1. Сильвестров Д.С. Программное обеспечение прикладной статистики. М., 1988, I-240.
2. Tiit, E.-M. Testing algorithms and programs. Int. J. of Pattern Recognition, 1989, v. 3, No 1. 121-133.

**МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ
ЭКОНОМИКИ ОТРАСЛИ ИНФОРМАЦИОННО-
ВЫЧИСЛИТЕЛЬНОГО ОБСЛУЖИВАНИЯ****А.Н.Гладун, Б.М.Кац**

Информационно-вычислительное обслуживание (ИВО) как отрасль народного хозяйства является новым видом общественного разделения труда. Промышленная обработка данных включает проектирование, производство и промышленную эксплуатацию информационно-вычислительных ресурсов.

Многомерный статистический анализ ИВО рассматривается в аспектах экзогенности и эндогенности системы (макро и микроэкономический анализ), классификации объектов наблюдения и исследования связи.

Макроэкономический анализ ИВО заключается в исследовании влияния информатизации на различные стороны деятельности общества и составляющие народнохозяйственного комплекса, в описании экономических моделей функционирования отрасли. Микроэкономический анализ заключается в исследовании хозяйственной деятельности отдельных вычислительных предприятий, групп предприятий, региона.

Анализ ИВО как объекта управления основывается на системном подходе к изучению общественных явлений и процессов. При исследовании проанализирована система обобщающих показателей, характеризующих уровень и эффективность использования различных видов ресурсов и отражающих специфические особенности индустрии информатики: инфраструктурность, наукоемкость, невещественность продукции, высокий уровень квалификации трудовых ресурсов, безсырьевое производство.

Информационной базой анализа ИВО являются данные статистического учета результатов хозяйственной деятельности вычислительных предприятий, представленные многомерной матрицей отношений объект-признаки. Предварительный статистический анализ позволил выделить для оценки хозяйственной деятельности вычислительных предприятий следующие показатели:

прибыль, объем и структура реализованных услуг, производительность труда, фондоотдача, себестоимость, рентабельность, структура парка вычислительной техники и другие.

Анализ экономики ИВО следует проводить как по всей совокупности предприятий, так и по отдельным типологическим группам ввиду существенного различия предприятий ИВО по своим характеристикам.

В исследовании для разбиения совокупности на группы использовались методы таксономии. Для выбора оптимального варианта разбиения использована процедура дискриминантного анализа. Имея несколько вариантов разбиения совокупности объектов на группы методами таксономии, необходимо построить дискриминантные модели для каждого из вариантов. Зная, к какой группе принадлежит объект и в какую он попал после дискриминации, можно судить о надежности дискриминантной модели. Критерием выбора варианта разбиения является максимальное значение надежности соответствующей дискриминантной модели.

Классификация вычислительных предприятий проведена на пространстве факторов фиксированной размерности, характеризующих использование информационно-вычислительных ресурсов.

Для оценки связи между группами наблюдаемых переменных использован канонический корреляционный анализ. Расчитаны корреляционные матрицы линейных связей между переменными по анализируемым вычислительным предприятиям в динамике. Согласование корреляционных структур во времени позволяет сделать вывод об устойчивости оценок корреляционных линейных связей.

Для качественной оценки резервов повышения эффективности ИВО проведено статистическое исследование связи и описаны регрессионные модели механизма образования прибыли в зависимости от показателей ресурсов в разрезе выделенных групп и в целом по совокупности. Получены результаты влияния на прибыль вычислительных предприятий объема работ, стоимости основных фондов, годового вычислительного потенциала, заработной платы и производительности труда.

ЭМПИРИЧЕСКИЕ ПАРАМЕТРЫ РАСПРЕДЕЛЕНИЯ В РЕАЛЬНОМ ПОТОКЕ СТАТИСТИЧЕСКИХ ДАННЫХ

Л.М.Тоодинг

1. Результаты наших исследований структуры реального потока задач статистического анализа данных свидетельствуют о вариативности статистических свойств исходных данных /1/. Ниже приводится характеристика эмпирического распределения признака, с тем чтобы дать представление о реальных предпосылках применения статистических методов. Основой анализа служили два информационных массива, составленные на базе архива статистических данных в Лаборатории прикладной математики ТГУ: 1) в целях характеристики формы кривой плотности распределения (данные о 2500 признаках-выборках), 2) для эмпирической характеристики параметров распределения и для сравнения различных оценок математического ожидания (700 признаков-выборок). Данные о признаках, среди которых были размах, стандартное отклонение, коэффициенты асимметрии и эксцесса, арифметическое среднее, медиана, мода, робастные оценки среднего значения и др., были получены и обработаны при помощи статистических пакетов, внедренных в ВЦ ТГУ.

2. Анализ этих информационных массивов показал, что предположение о нормальности распределения генеральной совокупности, являющееся основой классического статистического анализа, часто не выполняется: в 92% случаев пришлось отвергать гипотезу о согласии эмпирического распределения с нормальным распределением. В среднем наблюдается положительная асимметрия (коэффициент асимметрии: 0.72 ± 0.14 , интерквартильная область охватывает отрезок $[-0.1, 1.0]$). Коэффициент эксцесса также в среднем положительный, но распределен несимметрично: 60% из распределений имеет отрицательный эксцесс.

В ППП СТЕЛЛА, разработанном в нашей лаборатории, введена априорная типология признаков, с тем чтобы автоматически варьировать метод обработки данных. При этом различаются количественные показатели (тип R, 1/3 из признаков), целочисленные результаты подсчета (A, 1/10), качественные упорядоченные признаки (B, 1/2) и номинальные признаки (C, 1/10). Оказа-

лось /2/, что типология, введенная по предметному критерию (тип шкалы измерения), определяет четко разделяющиеся по статистическим свойствам классы переменных. На рис.1 изображены с 95%-ыми доверительными границами арифметические средние коэффициентов асимметрии и эксцесса и разности квартилей (в единицах стандартного отклонения).

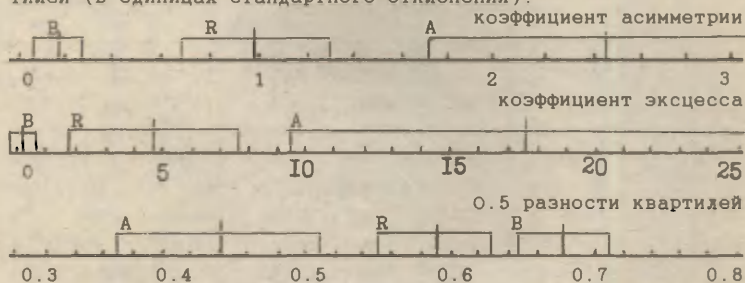


Рис.1.

3. Исходя из разнообразной статистической природы реальных данных, можно предполагать вариативность качеств различных статистических оценок. Нами были изучены более часто встречаемые в стандартных пакетах оценки центра распределения: мода, медиана, арифметическое среднее и три робастных средних, вычисляемые в пакете BMDP. Так как мы ограничивались эмпирическим подходом, где истинное среднее значение является неизвестным, то не было возможности характеризовать точность оценки, но удалось изучать их расстояния /2/. Мерой близости служили коррелятивное расстояние и уровень взаимного прогноза вариативности (квадрат корреляционного коэффициента). Оказалось, например, что эмпирическая мода имеет в среднем более различающееся от других оценок значение, а медиана немного лучше согласуется с робастными оценками, чем арифметическое среднее. Выявилась зависимость уровня связи между оценками от типа признака, причем наименьшее расхождение в оценках наблюдается у А-признаков.

Л и т е р а т у р а

1. Тоодинг Л.М. Описание структуры реального потока задач анализа данных. - Труды ВЦ ТГУ, 48, 1981, с.50-66.
2. Тоодинг Л.М. Изучение оценок среднего значения на базе реального потока задач. - Труды ВЦ ТГУ, 56, 1988, с.88-101.

СООТНОШЕНИЕ КАЧЕСТВА И ЦЕНЫ ПРИКЛАДНОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Н.А.Цибель

При установлении цены на пакет прикладных программ для персональной ЭВМ фирма-производитель исходит не столько из качества продукта, сколько из общего объема прибыли, которую она хочет получить от его реализации. Общий объем получаемой прибыли зависит от размеров рынка сбыта. Таким образом, цена на коммерческий программный продукт в значительной степени определяется возможностями его продажи. Можно продавать очень хороший пакет по сравнительно низкой цене, но большому числу пользователей и получить необходимую прибыль. У пакета низкого качества число покупателей ограничено, и чтобы получить нужную прибыль, он должен продаваться по сравнительно высокой цене.

Этого можно добиться с помощью рекламы, благодаря сложившимся связям фирмы с клиентами или тому, что определенную группу пользователей могут интересовать лишь отдельные, уникальные возможности пакета при плохой реализации или отсутствии большинства стандартных процедур. Кроме того, пользователь может купить довольно плохой пакет по высокой цене только потому, что он разработан для ЭВМ определенного типа, имеющейся у покупателя. Но этот случай можно не рассматривать, так как и у пакетов, предназначенных для машин одного класса, непосредственная зависимость между качеством и ценой не обнаруживается.

Американская фирма *National Software Testing Laboratories, Inc.* (Национальные Лаборатории по Тестированию Программ) ежегодно проводит рейтинг коммерческих пакетов для IBM-совместимых персональных компьютеров. В [1] публикуются оценки определенных классов пакетов по 10-балльной шкале: текстовых обработчиков; реляционных баз данных; графических пакетов; интегрированных пакетов многофункционального назначения и других. Никакой зависимости между ценами и баллами нет, причем сомневаться

в компетентности тестирующих и методике не приходится. Тестирование проводится по тщательно разработанной схеме с привлечением квалифицированных и неквалифицированных пользователей.

В [2] приводятся данные о количестве пользователей отдельных классов пакетов. Под пользователями понимаются организации или лица, закупившие лицензию на программный продукт или приобретшие копию программы и пользующиеся услугами по сопровождению со стороны фирмы-изготовителя. К сожалению, не для всех пакетов указано число пользователей и, кроме того, версии пакетов, представленные в [2], отличаются от рассмотренных в [1]. Поэтому, сопоставляя цены с баллами из [1] и объемом продаж из [2], можно лишь грубо оценить степень влияния последних на цену. Если в рамках одного класса сравнивать попарно пакеты разного качества, то в большинстве случаев оказывается, что цена пакета с более низким баллом выше, чем с более высоким, зато число пользователей меньше.

Рейтинг пакетов по прикладному статистическому анализу (ППСА) для персональных компьютеров в [1] не публикуется. Но если за качество ППСА принять число реализованных в нем статистических разделов и технических возможностей (это, пожалуй, единственное, что можно сделать на базе имеющейся информации), то из таблицы, приведенной в [3], видно, что связь между ценой и "универсальностью" (насыщенностью) пакета отсутствует. Однако данные о числе пользователей ППСА (см. [2]) так же не позволяют сделать вывод о влиянии этой величины на цену пакета.

Л и т е р а т у р а

1. *Software Digest Ratings Report*. 1984, vol. I; 1987, vol. 4; 1988, vol. 5.
2. *Datapro Directory of Microcomputer Software*. 1989, vol. 2, 3. McGraw-Hill, Inc. USA.
3. *PC-week*. 1988, vol. 5, № 26.

ПРИМЕНЕНИЕ МЕТОДОВ МНОГОМЕРНОГО АНАЛИЗА
В ИЗУЧЕНИИ ЭФФЕКТИВНОСТИ МАШИНОСТРОЕНИЯ БССР

В.А.Адамчик

Количественное выражение экономической эффективности производства остается одной из сложнейших теоретических и практических задач экономической науки. И на основе системы взаимосвязанных экономических показателей, и путём выделения из их числа одного главного, и при помощи конструирования нового интегрального показателя недостаточно и зачастую невозможно оценить эффективность. В этой ситуации определённый интерес представляет применение методов многомерного статистического анализа, два из которых – таксономии и главных факторов – были использованы при анализе эффективности машиностроения Белорусской ССР.

Каждая из одиннадцати отраслей машиностроения за годы 9, 10 и 11 пятилеток характеризовалась 50 показателями, которые были условно разделены на результирующие показатели эффективности и показатели условий производства (размер предприятия, основные производственные фонды, трудовые ресурсы, технический уровень производства, организация производственного процесса, экономическое стимулирование, социальное развитие). Расчёты выполнялись по методу выделения таксонов сложной формы. Анализ показал, что между обеими классификациями нет явно выраженной взаимосвязи: группировки в пространствах факторов и результирующих показателей не совпадают, то есть отрасли с лучшими характеристиками производства не являются самыми эффективными.

Задача выяснения влияния условий производства в отраслях машиностроения на эффективность решалась при помощи метода факторного анализа. С целью сокращения объёма информации и дублирования из 50 показателей были отобраны только 33. Оценки факторных нагрузок определялись, исходя из собственных значений и собственных векторов редуцированной корреляционной матрицы, рассчитанной по стандартизированным исходным данным. В дальнейшем расчёт велся по трём

факторам, для содержательной интерпретации которых проведено косоугольное преобразование факторного решения до получения простой структуры. Первый фактор объясняет 40,6% дисперсии всех показателей, включенных в модель, второй – 33,8%, третий – 25,6%. Определение факторных нагрузок проводилось по уравнениям регрессии, связывающим факторы с наблюдаемыми переменными. Мерой качества этой оценки служит коэффициент множественной детерминации, показывающий, какая часть дисперсии фактора обусловлена переменными, а какая остается необъясненной. Для трёх факторов этот коэффициент равен соответственно 88,5; 86,1 и 93,7%, что превышает критическое табличное значение. Следовательно, гипотеза о значимости коэффициента не отклоняется и факторам можно давать экономическую интерпретацию.

После анализа нагрузок первый фактор охарактеризован как размер производства, второй как структура затрат на выпускаемую продукцию, третий как технический уровень производства и труда в отрасли. Таким образом, факторный анализ позволил уточнить взаимосвязи изучаемых показателей и выделить только три группы переменных, каждую из которых можно интерпретировать и количественно выразить одной величиной – общим фактором.

Итак, в результате сформировались группы отраслей с высокой эффективностью – автомобильная и подшипниковая промышленность, тракторное и сельскохозяйственное машиностроение, приборостроение, со средней эффективностью – тяжелое, энергетическое и транспортное, строительно-дорожное и коммунальное, для легкой и пищевой промышленности машиностроение, с низкой эффективностью – станкостроение и инструментальная промышленность, электротехническая промышленность, химическое и нефтехимическое машиностроение.

В заключение следует отметить, что к недостаткам метода многомерного анализа, не позволяющим применять его достаточно широко, следует отнести неразработанность вопросов обеспечения сопоставимости во времени, а также распределения закономерностей, присущих всей совокупности, на отдельные единицы этой совокупности.

МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ В УГОЛЬНОЙ ПРОМЫШЛЕННОСТИ

Е.Н.Батугина

В терминах многомерного статистического анализа угольные разрезы представляют собой совокупность объектов, характеризующихся набором факторов. Этот набор включает в себя множество горно-геологических, горнотехнических, организационных, социально-экономических и климатических факторов. В качестве показателей, характеризующих результат работы горных предприятий выбираются следующие: производительность труда рабочего на добыче и на вскрыше и себестоимость тонны угля и кубометра вскрыши.

Для определения взаимосвязи между экономическими показателями и факторами, оказывающими на них влияние, строятся типологии угольных разрезов в пространстве экономических показателей и в пространстве факторов. Для классификации взято 70 разрезов Минуглепрома СССР. С использованием графо-аналитического способа многомерной классификации объектов по экономическим показателям было выделено три типологии (таксона) разрезов, объединенных по значениям показателей. В первый таксон (29 разрезов) попали такие разрезы, у которых высокая производительность труда на добыче, но низкая на вскрыше, что отразилось и на значениях себестоимости. Разрезы второго таксона (22 предприятия) имеют высокие значения показателей производительности труда на добыче и на вскрыше и относительно низкие величины себестоимости, т.е. с позиций эффективности эти предприятия более стабильны. Для предприятий третьего таксона (19 разрезов) характерны наиболее худшие значения экономических показателей.

Для построения типологий в пространстве влияющих факторов статистическим анализом отобраны наиболее информативные, такие как коэффициент вскрыши, обводненность месторождения, мощность пласта, угол падения, вид транспорта на добыче и на вскрыше, тип экономического микрорайона, комплексный климатический показатель. Эти факторы были предварительно переведены в безразмерные координаты, что позволило соизмерить объекты, описанные разнотипными факторами. В пространстве этих факторов выделено три таксона угольных разрезов. Разрезы, отнесенные к первому таксону (19 разрезов), характеризуются в основном простыми горно-геологическими

условиями: мощность пластов до 10 и более метров, падение их горизонтальное и пологое. На вскрыше в основном применяется бестранспортная система. Для разрезов второго таксона (16 разрезов) характерны значительная мощность пластов, угол падения пластов от наклонного до крутого, все разрезы расположены в климатических зонах с континентальным климатом, малообводненные. Представителями третьего таксона (35 разрезов) стали почти все объекты, которые разрабатывают по несколько пластов значительной мощности и крутого падения, коэффициент вскрыши до $10 \text{ м}^3/\text{т}$, с климатической точки зрения район повышенной сложности.

Геснота связи между классификациями угольных разрезов в пространстве экономических показателей и в пространстве факторов оценена коэффициентом взаимной сопряженности Чупрова, который оказался равным 0,405.

Построенные типологии угольных разрезов могут являться основой при оценке технического уровня предприятий, а также при разработке мероприятий по повышению эффективности работы разрезов отдельных типологий.

Влияние факторов, измеренных в разных шкалах на экономические показатели, оценивается на основе элементов распознавания образов. С этой целью из всей совокупности разрезов выделены объекты, у которых значение экономического показателя выше среднотраслевого (класс А) и такие разрезы, у которых значение показателя ниже среднотраслевых. Наличие двух противоположных классов объектов позволяет определить, как и чем объекты одного класса отличаются от объектов другого класса и построить решающее правило прогноза (распознавания) значений экономического показателя. Например, для прогноза значения фондоотдачи на угольных разрезах построено следующее решающее правило:

$$\begin{aligned} \text{если } P_1 &\leq 1,97, \text{ то } P \in A = 0, \\ \text{если } P_1 &\geq 2,45, \text{ то } P \in A = 1,0, \end{aligned}$$

где P_1 — функция распознавания.

Построенные модели прогноза по различным экономическим показателям для угольных разрезов отрасли позволяют определять вероятность принадлежности разреза к классу с высокими экономическими показателями и управлять объектом, если он попадает в нежелательный класс. Для этого достаточно изменить значения управляемых факторов так, чтобы объект оказался в желаемом классе.

ИССЛЕДОВАНИЕ ЗАКОНОМЕРНОСТЕЙ РАЗВИТИЯ
РЕГИОНАЛЬНЫХ СИСТЕМ ОБЛАСТНОГО РАНГА
КЛАССИФИКАЦИОННО-ТИПОЛОГИЧЕСКИМИ МЕТОДАМИ

И.Б.Береснева

Приводятся результаты исследования структурной динамики регионов РСФСР областного ранга. Информация о рассматриваемых объектах представлялась системой значений выраженности отраслевых функций [1] по данным на 1965, 1970, 1975, 1980 и 1984 годы. Модельное пространство исследуемого процесса строилось методами многомерного шкалирования в соответствии с [1.]

При выделении и анализе типологических группировок изучаемых объектов в модельном пространстве возникает проблема совместного использования методов многомерного шкалирования и классификации [2,3]. Приводятся результаты исследования этой проблемы в рамках функциональных моделей многомерного шкалирования и предлагаются методы корректного сочетания многомерного шкалирования и классификации при решении прикладных задач.

Программное и методическое обеспечение использовалось в рамках системы анализа социально-экономических данных (САСЭД) ИСЭП АН СССР. Для классификации объектов применялись методы построения минимальной ξ -сети, а также алгоритмы, связанные с минимизацией критериев дисперсионного типа [4].

Результатом применения предложенного классификационно-типологического подхода является классификация исходного множества объектов, в которой каждому классу соответствует точка модельного пространства (агрегированного описания структуры рассматриваемых объектов). Построенная классификация является основой для выделения типологических группировок объектов. С её помощью описывается и структура временных переходов объектов из класса в класс.

Приведены результаты разбиения общего процесса структурного развития на две компоненты: первая обусловлена инерционными законами развития, вторая - воздействием внешних "возмущений" структуры системы.

В докладе приводятся конкретные результаты исследования областных региональных систем, полученные в рамках предложенного подхода.

Л и т е р а т у р а

1. Крупный социалистический город: структурные аспекты развития /Отв. ред. Н.Т.Агафонов.-Л.:Наука, 1987. - 142 с.
2. Крускал Дж. Взаимосвязь между многомерным шкалированием и кластер-анализом // Классификация и кластер.-М.:Мир, 1980. - С. 20-41.
3. Shepard R.N. Representation of structure in similiary data: Problems and prospects // Psychometrika. - 1974. - v.39 - P. 373-422.
4. Береснева И.В., Маслякова Т.В., Минина Т.Р. и др. Применение методов квазидинамического программирования в задачах снижения размерности.-Л.,1986. - 70 с.

ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ ППП ОЛИМП И СКИФ

В.В.Богачев, И.П.Колпаков

Задача анализа и прогнозирования временных рядов экономических показателей является составным элементом комплексного анализа хозяйственной деятельности различных структурных звеньев экономической системы и должна решаться на основе единой информационной модели экономического объекта. Под единой информационной моделью здесь понимается совокупность экономических показателей, определяющих различные аспекты деятельности объекта.

Исходя из этого система прогнозирования должна включать средства ведения информационной модели объекта и средства прогнозирования временных рядов экономических показателей.

Предлагаемый программный комплекс (ПК) включает систему организации и ведения информационной модели экономического объекта СКИФ и систему анализа, моделирования и прогнозирования временных рядов ОЛИМП.

ОЛИМП реализует следующие методы прогнозирования сезонных и несезонных временных рядов:

- адаптивные методы Брауна, Хольта и Хольта-Уинтерса;
- метод параметрических моделей Бокса-Дженкинса;
- метод обобщенной линейной параметрической модели.

Для поиска оптимальных параметров используется диалоговая система нелинейной оптимизации.

СКИФ обеспечивает организацию, ведение базы данных сетевого типа, доступ к информации по сложным критериям выбора и позволяет в рамках заданной модели объекта формировать и выдавать различные совокупности экономических показателей, необходимых для работы ОЛИМП. Информационный интерфейс между ОЛИМПом и базой данных реализуется СКИФом автоматически.

ПК функционирует на ПЭВМ класса ЕС-1840 под управлением операционной системы MS DOS.

РАЗРАБОТКА СИСТЕМЫ МОДЕЛЕЙ МИРОХОЗЯЙСТВЕННЫХ СВЯЗЕЙ С РЕАЛИЗАЦИЕЙ АДАПТИВНОГО ПОДХОДА

И.М.Денисова, О.А.Андрюшкевич

Моделирование социально-экономических процессов опирается на богатый арсенал математических методов, среди которых регрессионный анализ занимает значительное место. Практика использования регрессионного анализа показала, что он имеет ряд существенных преимуществ по сравнению с другими методами, связанных не только с глубиной и широтой анализа исследуемых процессов, но и относительной простотой его реализации на ЭВМ [1]. Кроме того, он хорошо зарекомендовал себя и в прогнозных расчетах. Его некоторая ограниченность использования для прогноза (имеется в виду постоянство во времени обнаруженных в результате анализа взаимосвязей между переменными и их вклада в динамику процесса) может быть преодолена с помощью широко известного сценарного подхода, а также путем динамизации коэффициентов регрессионного уравнения. Последний подход особенно хорошо зарекомендовал себя в случае сложных экономических процессов, динамика которых существенно меняется по времени.

Однако в ряде случаев целесообразно использовать метод адаптации. Применение адаптивного подхода позволяет построить модель с переменными структурными коэффициентами, которые дают представление не только о характере зависимости в каждый момент времени, но и строить прогнозы на основе ценной экономической информации, содержащейся в динамических рядах коэффициентов регрессионных уравнений.

Математическая постановка задачи выглядит следующим образом [2, с. 215]. Найти минимум функции вида:

$$I_t(\omega_t) = \sum_{i=1}^M \varphi_{it}^2 (F_{it} - f_{it})^2 = \sum_{i=1}^M (F_{it} - \sum_{j=0}^N \omega_{ij}^t \varphi_{jt})^2 = \min \quad (1)$$

F_{it} - фактическое значение экономических показателей;

f_{it} - расчетное значение экономических показателей;

N - число факторов в уравнениях регрессии;

M - число уравнений регрессии;

$\varphi_{jt} = 1$ при $j = 0$.

Для адаптации коэффициентов α_{jt} используется метод наискорейшего спуска по градиенту функционала [4, с. 216]. После соответствующих преобразований получим формулу (2), по которой на каждом шаге будем осуществлять обновление коэффициентов:

$$A_{jt} = A_{jt} + \frac{\gamma}{\sum_{j=1}^N \varphi_{jt}^2} (F_{jt} - f_{jt}) \cdot \varphi_{jt} \quad (2)$$

A_{jt} - вектор старых коэффициентов уравнения регрессии;

A_{jt} - вектор новых коэффициентов уравнения регрессии;

γ - коэффициент ($\gamma > 0$).

Коэффициент γ определяет реакцию модели на полученную текущую ошибку и корректирует коэффициенты регрессии в нужном направлении. Оптимальные значения γ определяются в процессе обучения модели [3, с. 217].

Практическая реализация данного метода была проведена при изучении валютно-финансовых проблем капиталистических стран. Так при анализе и моделировании валютно-финансового механизма Франции, ФРГ и Великобритании был использован адаптационный алгоритм. Модели имеют соответственно 13, 11 и 11 уравнений и 1 тождество. Уравнения оценивались на периоде 1960-1984. По одной из рассматриваемых стран было проведено сравнение результатов адаптационного подхода с расчетами, полученными на основе использования рекурсивной системы регрессионных уравнений. Результаты анализа и полученные пост-прогнозы убедительно показали преимущества адаптационного подхода для исследования процессов, имеющих неустойчивую динамику и подверженных сильному влиянию конъюнктурных факторов.

Л и т е р а т у р а

1. О. А. Андрушкевич, И. М. Денисова Изучение взаимосвязей социально-экономического развития региона с помощью эконометрической модели. Тезисы докладов. Тарту, 1985.
2. Левицкий Е. М. Адаптивные эконометрические модели, Новосибирск, Наука, 1981.
3. Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования. М., Статистика, 1979.

ОЦЕНИВАНИЕ ТРЕНДА ПАРАМЕТРОВ В ЭКОНОМЕТРИЧЕСКИХ ЗАВИСИМОСТЯХ

Н.И.Киселев

Предположим, что имеется ряд наблюдений за n лет над результатом деятельности y экономической системы и факторами производства x_1, x_2, \dots, x_p . Для описания зависимостей между ними выбрана некоторая функция, параметры которой неизвестны и, более того, имеют дрейф во времени. Задача состоит в оценке трендов этих параметров на основе фактических данных с привлечением экономических соображений.

Рассмотрим подход на примере производственной функции типа Кобба-Дугласа

$$y = A_0 x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot \dots \cdot x_p^{\alpha_p}$$

Прямой путь в "отслеживании" динамики коэффициентов $\alpha_1, \dots, \alpha_p$ состоит в разбиении общего числа n наблюдений на подвыборки последовательных наблюдений объемом $p+1$, т.е. устраивается скользящий просмотр от первой подвыборки, состоящей из первых $p+1$ наблюдений $1, 2, \dots, p$ до последней - из наблюдений $n-p, n-p+1, \dots, n$. Для каждой подвыборки находим оценки коэффициентов $\alpha_1, \dots, \alpha_p$ путем решения системы линейных уравнений. Таким образом, для каждого коэффициента получаем динамический ряд, состоящий из $n-p+1$ значений. Однако, в силу высокой мультиколлинеарности факторов производства оценки коэффициентов крайне неустойчивы и изучение их тренда по полученным временным рядам приводит к неудовлетворительным результатам.

Один из возможных путей состоит в переходе от исходных факторов к слабо коррелированным инструментальным переменным. Для их получения используется метод главных компонент следующим образом. Вначале осуществляется переход к логарифмическому масштабу факторов производства. Для

преобразованной матрицы данных определяются главные компоненты, от значений которых берется экспонента. В результате получаем матрицу наблюдений набора $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p$ инструментальных переменных, где i -я переменная имеет вид

$$\bar{x}_i = c_{0i} \cdot x_1^{c_{1i}} \cdot x_2^{c_{2i}} \cdot \dots \cdot x_p^{c_{pi}}$$

и, соответственно, производственная функция

$$y = \beta_0 \cdot \bar{x}_1^{\beta_1} \cdot \bar{x}_2^{\beta_2} \cdot \dots \cdot \bar{x}_p^{\beta_p}$$

Здесь (c_{1i}, \dots, c_{pi}) представляет i -ый собственный вектор соответствующей ковариационной матрицы. Учитывая выборочную природу оценок (c_{1i}, \dots, c_{pi}) в конкретных задачах целесообразно некоторое их изменение так, чтобы инструментальные переменные приобрели более ясное экономическое содержание.

Используя упомянутую последовательность подвыборок, но уже с переменными $y, \bar{x}_1, \dots, \bar{x}_p$ получаем динамические ряды коэффициентов β_1, \dots, β_p , которые сравнительно с $\alpha_1, \dots, \alpha_p$ устойчивы и, как правило, имеют экономически интерпретируемый тренд.

Пример. Рассматривается совокупный объем производства y потребкооперации и факторы производства: x_1 — стоимость основных фондов, x_2 — фонд оплаты труда, x_3 — оборотные средства за 1976–1987 годы.

Первая главная компонента, вычисленная для матрицы наблюдений факторов в логарифмическом масштабе объясняет 96,3% и две компоненты — 99,5% общей изменчивости.

Первая и вторая инструментальные переменные имеют вид $\bar{x}_1 = x_1^{0.66} \cdot x_2^{0.42} \cdot x_3^{0.51}$ и $\bar{x}_2 = x_1^{-0.58} \cdot x_2^{-0.21} \cdot x_3^{0.78}$. Переменная \bar{x}_1 допускает интерпретацию как масштабный фактор производства, а \bar{x}_2 трактуется как фактор, учитывающий "внутреннее" соотношение факторов производства, т.к. она инвариантна к масштабу производства.

Выражения для трендов коэффициентов β_1 и β_2 имеют вид

$$\beta_1(t) = 1.5 - 0.08t \quad \text{и} \quad \beta_2(t) = 18 + 16.1t$$

На содержательном уровне имеет место вывод: коэффициент эластичности совокупного объема производства по масштабному фактору уменьшается во времени, тогда как важность структурного фактора увеличивается.

ПРОГНОЗИРОВАНИЕ СПРОСА НА ОСОБО МОДНЫЕ ИЗДЕЛИЯ В ЛЕГКОЙ ПРОМЫШЛЕННОСТИ

А. Б. Колбасова

В легкой промышленности спрос на товары зависит от многих быстро меняющихся факторов. Нами разрабатывается методика определения кратковременного спроса на особо модные изделия, реализуемые по договорным ценам. Исследования проводятся в одном из фирменных магазинов МПШО "Москва" "Мужские сорочки".

При существующих направлениях моды можно выделить несколько основных факторов спроса на модель. Однако их действие варьирует, поскольку оно связано с социальными, демографическими, психологическими характеристиками покупателей. По результатам одного исследования, прогноз имеет смысл делать не менее чем на 20 дней, тогда с 90% вероятностью можно утверждать, что доля желающих купить модель не будет отклоняться от средней более чем на 19%.

Первый фактор - поток покупателей. На него оказывают влияние не только сезонные колебания, но и тренд, который определяется конкурентоспособностью товаров данного магазина на региональном рынке и степенью удовлетворенности покупателей качеством этих товаров. Измерение второй составляющей тренда требует проведения такого опроса, при котором можно было бы выяснить: 1) значимость для анкетированного таких критериев покупки, как цена, качество изготовления модели, состав ткани, функциональные качества (практичность, торжественность, применяемость во многих

сферах), модность, эмоционально-эстетические свойства; 2) степень удовлетворенности покупателя этими параметрами, определяемой по шкале. Эти данные целесообразно использовать для сегментации покупателей и изучения на этой основе, привлечет ли модель новую часть рынка. Разумеется, для выделения типов покупателей нужно расширить набор изучаемых признаков. Параллельно следует провести классификацию моделей по внешним признакам и выделить типичные группы взаимозаменяемых товаров.

Эта задача сопряжена с проблемой измерения еще одного фактора спроса - наличия в продаже конкурирующих моделей. Заметим, что среди них есть и модели-лидеры, спрос на которые почти не меняется с внедрением новых моделей. Выделение типов моделей позволит по таким отдельным элементам, как, например, силуэт, форма воротника, расцветка, предсказать для данной модели ее товаров-конкурентов. Для оценки этого фактора спроса возможно применение других методов: 1. В первой выборке покупателей респонденту предлагается на рассмотрение одна модель, а в другой опрос по этой модели проводится, когда она представлена на фоне двух других. Здесь отражается проблема измерения влияния на спрос дефицитности, отсутствия выбора товаров. 2. Проведен опрос покупателей для выяснения того, какие модели они бы купили, если бы отсутствовала наиболее им понравившаяся. 3. Анализ фактических продаж моделей до и после выставки-продажи новой модели.

Третий фактор - цена, которую согласовывают между собой торговое и промышленное предприятия. Для его измерения проводится устное анкетирование покупателей. Основная цель опроса - узнать, по какой максимальной цене респондент согласится купить изучаемую модель. Затем для разных моделей проводится аппроксимация эмпирических кривых распределения ожидаемого спроса в зависимости от цены изделия. Оказалось, что для каждой модели достаточно ввести в функцию лишь медианное значение цены, которое называют 50% желающих купить модель.

СТАТИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ АДЕКВАТНОСТИ ОЦЕНКИ НАУЧНО-ТЕХНИЧЕСКОГО УРОВНЯ АСУТП

И.Н.Минскер

В настоящее время большое внимание уделяется проблемам повышения научно-технического уровня (НТУ) и качества продукции. Методика оценки НТУ автоматизированных систем управления технологическими процессами (НТУ АСУТП) была разработана в Центральном НИИ комплексной автоматизации в 1985 году [1]. Процедура оценки НТУ АСУТП, согласно [1] заключается в следующем:

1. Формируется перечень свойств, определяющих качество АСУ и выбираются аналоги: лучшие и средние АСУ данного типа.

2. На основе сравнения с аналогами оценивается уровень каждого свойства АСУ в четырехбалльной ранговой шкале: А - выше мирового уровня; В - соответствует мировому уровню; С - соответствует среднему уровню; Д - ниже среднего уровня.

3. Производится агрегирование ранговых показателей уровня методом бинарной композиции [2].

В 1987-1988г. методика прошла экспериментальную проверку при анализе качества ряда АСУТП, разрабатываемых в соответствии с комплексной программой научно-технического прогресса стран СЭВ.

Для проверки адекватности принятого в методике оценки НТУ алгоритма агрегирования, использующего экспертные матрицы свертки, был использован алгоритм построения агрегированных количественных показателей по массиву первичных ранговых показателей [3].

Результат анализа показал полное совпадение классификации АСУТП, полученной на основе НТУ, рассчитанного по методике [1] и классификации на основе количественного показателя, полученного по алгоритму [3]. Это совпадение свидетель-

ствует об обоснованности алгоритма свертки показателей в методике оценки НТУ и позволяет рекомендовать его к широкому применению.

Л и т е р а т у р а

1. Левин А.А., Минскер И.Н., Хвилевцкий Л.О. Методические вопросы оценки НТУ АСУТП. Приборы и системы управления, 1988 № II, 1-5.
2. Глотов В.А., Павеев В.В. Векторная стратификация. - М., Наука, 1984.
3. Баумштейн П.И. Один метод агрегирования ранговых признаков. //Тезисы докладов IV Всес.конф. "Применение многомерного статистического анализа в экономике и оценке качества продукции". - Тарту, 1989.

ЗАДАЧА О ВЫБОРЕ ТИПОПРЕДСТАВИТЕЛЕЙ В ОБУВНОЙ ПРОМЫШЛЕННОСТИ

В.Ольман

Очевидно, что изготавливать всю обувь по индивидуальным заказам — задача немислимая как экономически, так и организационно. Альтернативным подходом является разработка конечного числа обувных типопредставителей, среди которых потребитель и выбирает наиболее подходящую для себя обувь. Задача разработчика заключается в построении такого набора типопредставителей, чтобы минимизировать среднюю в некотором смысле неудовлетворенность пользователей обуви.

Математическая формулировка задачи выглядит следующим образом. Пусть требования каждого отдельного индивида к обуви описываются вектором $y \in R^k$, $k \geq 1$, (компоненты этого вектора заданы в ГОСТ 11373-75, ГОСТ 3927-75), причем известно вероятностное распределение F вектора требований пользователей. Кроме того пусть неудобство пользователя с требованиями y при ношении обуви, соответствующей требованию X , количественно определяется неотрицательной функцией потерь $L(y, X)$. Пусть заранее задано число возможных типопредставителей $m > 1$ (оно может быть определено, например, суммой, выделенной на разработку всех типопредставителей). Тогда, если разработаны типопредставители с параметрами x_1, x_2, \dots, x_m , то средняя неудовлетворенность пользователей определяется величиной

$$\Psi(x_1, x_2, \dots, x_m) = \sum_{i=1}^m \int_{Y_i} L(y, x_i) dF(y),$$

где Y_i — это множество тех пользователей, для которых типопредставитель x_i наиболее пригоден, т.е.

$$Y_i = \{y : L(y, x_i) \leq L(y, x_j), j = 1, 2, \dots, m\}.$$

Для минимизации функции $\Psi(x_1, x_2, \dots, x_m)$, т.е. для поиска оптимального набора типопредставителей, можно использовать либо метод Ллойда [1], который пригоден и в случае известной

функции $F(y)$ и когда у исследователя есть статистические наблюдения над случайной величиной y , либо дифференциальный подход [1], использующий точное знание функции $F(y)$.

В сообщении рассмотрен пример, когда $k=1$, т.е. требования индивида определяются одномерным параметром (размером или полнотой), а функция $L(y, x)$ имеет вид

$$L(y, x) = \begin{cases} \alpha(x-y), & x \geq y \\ \beta(y-x), & x < y, \end{cases}$$

причем $\beta \gg \alpha > 0$, так как неудовлетворенность потребителя "жмущей" обувью существенно больше, чем свободной.

При анализе описанной модели были использованы следующие математические результаты.

Лемма. Пусть $dF(y) = m(y)dy$, причем носитель F , т.е. множество $\mathcal{Y}_F = \{y: m(y) > 0\}$ является интервалом. Тогда для того, чтобы $(x_1^*, x_2^*, \dots, x_m^*)$ минимизировали бы функцию $\Psi(x_1, \dots, x_m)$ необходимо, чтобы

$$L(b_i^* - x_i^*) = L(x_{i+1}^* - b_i^*), \quad (I)$$

где $(b_i^*, b_{i+1}^*) = \mathcal{Y}_i$, $i = 0, 1, 2, \dots, m-1$, $b_0^* = -\infty$, $b_m^* = \infty$.

Теорема. Пусть $\ln m(y)$ является вогнутой функцией на \mathcal{Y}_F . Тогда система (I) имеет единственное решение т.е. функция $\Psi(x_1, x_2, \dots, x_m)$ одноэкстремальна.

Л и т е р а т у р а

1. S.P. Lloyd. Least squares quantization in PCM. IEEE Trans. Inform. Theory, vol. IT-28, pp.129-137, Mar. 1982.

МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ ПОКУПАТЕЛЬСКОЙ МОТИВАЦИИ ВЫБОРА ТОВАРА

И.В.Семеняк, Е.И.Березань

Одним из условий разрешения возникшего неантагонистического противоречия между сформированным спросом и качественной ущербностью рынка ТНП является информация о тенденции изменения спроса на товары определенного качества.

Покупатель выражает свое отношение одновременно ко многим сторонам качества товара. Сочетание избирательных признаков в значительной мере характеризует социально-психологический портрет покупателя. Возникает задача многомерной группировки по сочетанию альтернативных признаков, требующая теоретического решения и разработки практической методики. Применению формального метода многомерной группировки по сочетанию качественных признаков должна предшествовать неметрическая процедура, приводящая к метрическому результату. В современной научной литературе нет описаний решения поставленной задачи [1, 2].

Предложена методика распознавания образов по сочетанию качественных признаков, закодированных как "1" или "0". Матрица сочетаний альтернативных признаков с признаком-лидером, в связи с которым определяется система положительных альтернатив ($X_k = I$), создается по критерию логического умножения признака-лидера (x_j^i) и других признаков, вошедших в программу наблюдения (x_j^p):

$$X_{kj} = \begin{cases} \sum_{i=1}^n x_j^i x_j^p, & \text{с } x_j^i \wedge x_j^p = 1; \\ 0, & \text{с } x_j^i \wedge x_j^p = 0 \end{cases}$$

Итоговая матрица сочетаний признаков принимает вид:

$$y = \begin{pmatrix} \sum x_1 x_1 & \sum x_1 x_2 & \dots & \sum x_1 x_K \\ \sum x_2 x_1 & \sum x_2 x_2 & \dots & \sum x_2 x_K \\ \dots & \dots & \dots & \dots \\ \sum x_K x_1 & \sum x_K x_2 & \dots & \sum x_K x_K \end{pmatrix} \quad \begin{array}{l} \text{где } \sum x_K x_K - \text{частота} \\ \text{альтернативного} \\ \text{признака } x_K^i. \end{array}$$

Имея диагональную матрицу частоты сочетаний признаков, можно выполнить многомерный анализ с целью агрегирования признаков и расчленения совокупности на типы. В исследовании мнений 2 тысяч покупателей электронной радиоаппаратуры г. Харькова агрегирование выполнено в главные компоненты U_m методом факторного анализа.

Вся совокупность обследованных покупателей разделена на две основные общности, слабо коррелирующие между собой по мотивам предпочтения изделий. По существу признаков первая общность - новаторы, вторая - рациональные покупатели. Вычленяется промежуточная общность рациональных новаторов. Установлена структура совокупности.

Результаты анализа использованы Госкомстатом УССР, включены в ГАСНТИ.

Разработанная и описанная методика дифференциации статистической совокупности по сочетанию альтернативных признаков, компонентного анализа диагональной матрицы сочетаний может быть использована в различных социально-экономических исследованиях, связанных с изучением покупательского спроса и общественной оценкой качества ТНП.

Л и т е р а т у р а

1. Миркин Б.Г. Группировки в социально-экономических исследованиях. - М.: Финансы и статистика, 1985.
2. Жамбю М. Иерархический кластер-анализ и соответствия. Пер.с фр. - М.: Финансы и статистика, 1988.

ПРИМЕНЕНИЕ МЕТОДОВ КЛАСТЕРИЗАЦИИ И РАНГОВОЙ КОРРЕЛЯЦИИ В ИЗУЧЕНИИ ТЕНДЕНЦИЙ РАЗВИТИЯ ПРОМЫШЛЕННЫХ ПРЕДПРИЯТИЙ

И.В.Семеняк, П.А.Иващенко

Предложен подход к проблеме комплексной оценки сдвигов в научно-техническом и социально-экономическом развитии предприятий при реализации радикальной экономической реформы.

Исходные данные представлены совокупностью предприятий $Q = \{q_i\}_{i=1}^N$, каждое из которых характеризуется системой признаков $\Pi = \{\pi_j\}_{j=1}^m$. Считается, что Π состоит из совокупностей Π^I - признаков порядковой шкалы и $\Pi^{\bar{I}}$ - признаков шкалы интервалов, результативного признака Π_1 : $\Pi = \Pi^I \cup \Pi^{\bar{I}} \cup \Pi_1$. Измерения значений признаков расположены в двух прямоугольных таблицах с N строками и m столбцами. Они описывают состояние предприятий соответственно "до" и "после" перехода на новые условия хозяйствования. Обсуждаются две актуальные задачи [1]: 1. Задача многомерных структурных группировок заключается в нахождении совокупности из трех кластеров $\{Q_1, Q_2, Q_3\}$ (передовые, средние и отстающие предприятия). 2. Задача "нащупывания" изменений, вносимых хозяйственной реформой по силе взаимовлияния признаков.

Авторами разработаны оригинальные методы решения этих задач применительно к конкретной совокупности предприятий легкой промышленности Харьковской области, перешедших в 1987 году на новые условия хозяйствования. Основной математический результат исследования - методика взвешенного кластерного анализа и метод экстремальных точек, экономико-статистический вывод - необходимость принятия управленческих решений по ускорению внедрения нововведений.

Л и т е р а т у р а

1. Мандель И.Д. Кластерный анализ. - М.: Финансы и статистика. 1988.

МОДЕЛИРОВАНИЕ ЛОНГИТУДНЫХ ДАННЫХ

С.Ю.Адамов, И.А.Герасимова

При анализе таблиц сопряженности, особенно в тех случаях, когда на пересечении i -й строки и j -го столбца стоит количество объектов, перешедших из i -го состояния в j -е (матриц обмена), может оказаться полезной модель квази-симметричности, предложенная А.Коссинсом [1].

$$f_{ij} = a_i b_j s_{ij}, \quad (1)$$

$i=1, \dots, n$; $j=1, \dots, n$. Здесь f_{ij} — оценка согласно модели элемента пхл исходной таблицы x_{ij} ; a_i , b_j и s_{ij} — параметры.

По x_{ij} трудно судить о близости i -го и j -го состояний: во-первых, в общем случае $x_{ij} \neq x_{ji}$, а во-вторых на число переходов кроме близости состояний влияют еще различия в маргинальных частотах исходного и конечного состояний. Модель (1) разделяет эти эффекты. Очевидно, что при увеличении a_i согласно модели увеличивается число объектов, покидающих i -е состояние, а при увеличении b_j — увеличивается число объектов, переходящих в j -е состояние. Поэтому они называются индексами отталкивания и привлекательности соответственно. Параметры s_{ij} характеризуют близость между состояниями.

Ввиду неоднозначности определения параметров предлагается дополнительное условие:

$$s_{ij} = s_{ji}, \quad \sum_i a_i = \sum_j b_j = 1$$

В [1] параметры f_{ij} оцениваются с помощью пакета "GLIM", затем матрица $\{f_{ij}\}$ симметризуется. Полученная симметричная матрица $\{s_{ij}\}$ преобразуется так, чтобы сумма элементов в каждой строке и в каждом столбце равнялась единице. Ниже предлагается алгоритм, позволяющий непосредственно находить оценки параметров, m — номер итерации. Процесс останавливается, когда на соседних шагах не

происходит существенных изменений оценок. Доказано, что этот алгоритм сходится и дает оценки максимального правдоподобия.

$$0) \alpha_t^{(0)} = \frac{x_t}{N}; s_{t,j}=1; t=1, \dots, n; j=1, \dots, n.$$

$$1) b_j^{(m)} = \frac{x_{\cdot j}}{\sum_t \alpha_t^{(m-1)} s_{t,j}} \text{ при } j=1, \dots, n.$$

$$2) \alpha_t^{(m)} = \frac{x_t}{\sum_j b_j^{(m-1)} s_{t,j}}; t=1, \dots, n.$$

$$3) s_{t,j}^{(m)} = \frac{x_{t,j} + x_{j,t}}{\alpha_t^{(m-1)} b_j^{(m-1)} + \alpha_j^{(m-1)} b_t^{(m-1)}}; t=1, \dots, n; j=1, \dots, n.$$

Модель квази-симметричности оказывается наиболее продуктивной, когда имеется возможность сравнивать значения параметров для различных групп или в различные моменты времени. Для этого предлагаются две модели:

$$f_{t,j,t} = \alpha_{tt} b_{jt} s_{t,j,t}, s_{t,j,t} = s_{j,t,t}; \sum_{j,t} \alpha_{tt} = \sum_{j,t} b_{jt} = 1 \quad (2)$$

$$f_{t,j,t} = \alpha_{tj} b_{jt} s_{t,j,t}, s_{t,j,t} = s_{j,t,t}; \sum_t \alpha_{tj} = \sum_j b_{jt} = 1 \quad (3)$$

$t=1, \dots, n; j=1, \dots, n; t=1, \dots, T; t \neq j.$

Предложенный алгоритм обобщается для оценки параметров каждой из этих моделей. Статистики согласия для моделей (2) и (3) имеют распределение χ^2 с $(2T-1)(n-1)(n-2)/2 - (T-1)$ степенями свободы [2].

ЛИТЕРАТУРА

- [1] H.Caussinus "Quelque points de vue sur l'analyse des tableaux d'echanges" Annales de L'INSEE N 22-23/1976.
- [2] H.Caussinus, C.Thelot "Note complementaire sur l'analyse statistique des migrations" Annales de L'INSEE N 22-23/1976.

МОДЕЛИРОВАНИЕ ВЗАИМОСВЯЗИ ПОКАЗАТЕЛЕЙ СОЦИАЛЬНОЙ АКТИВНОСТИ РАБОТНИКОВ АПК

В.А.Балаш, Н.Ф.Горчакова

В работе рассматриваются вопросы применения логлинейного анализа таблиц сопряженности для моделирования взаимосвязи показателей социальной активности. Характеристика уровня социальной активности осуществлялась при помощи показателей, построенных на основе данных анкетного обследования работников ряда сельскохозяйственных предприятий. Показатели строились по оценкам степени развития различных сторон активности самими работниками.

Логлинейный анализ основан на предположении о достаточной заполненности ячеек таблицы сопряженности. Прямое его применение к изучению различных взаимосвязей одновременно большого числа признаков невозможно, так как общее число ячеек такой таблицы обычно слишком велико. Поэтому поиск существенных связей проводился на основе анализа отдельных фрагментов многомерной таблицы сопряженности. Данный прием основывается на предположении, что связи высокого порядка, если они достаточно сильны, должны проявлять себя и в связях более низкого порядка, которые могут быть исследованы по фрагментам многомерной таблицы.

Предварительно, на основе матрицы парных коэффициентов, выделялись наиболее существенные связи между отдельными показателями, и строился граф существенных парных связей.

Для множества вариантов, полученных на основе такого графа, проводилось построение логлинейных моделей взаимосвязи показателей. Целенаправленный поиск существенных связей позволил сократить число вариантов расчетов, и в то же время выявить ряд существенных взаимодействий третьего и четвертого порядка. Рассмотрение указанных взаимодействий показало, что усиление активности в одной из форм не всегда сопровождается пропорциональным возрастанием активности в других формах, рост активности идет по отдельным направлениям.

О ПРИМЕНЕНИИ МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА В ИЗУЧЕНИИ ДЕМОГРАФИЧЕСКОЙ СИТУАЦИИ

Н.В.Высоцкая, О.В.Терещенко

Процессы и явления, происходящие в городах, взаимосвязаны друг с другом, в них определенным образом сочетаются частные и общие причины и следствия и проявляются случайность и необходимость. В связи с этим характерной чертой демографической ситуации в городах различной людности [4] является ее многомерность. Это свойство включает несколько элементов, каждый из которых характеризуется группой свойств, а каждое свойство в свою очередь описывается многими признаками.

Задача заключается в том, чтобы свести исходное многообразие изучаемых городов к сравнительно небольшому числу однородных различных типов, совокупность которых можно рассматривать как структуру городов изучаемого региона, проанализировать и описать каждый из полученных типов. Здесь речь идет об эмпирических типах, как устойчивых сочетаниях свойств объектов (городов), рассматриваемых в нескольких измерениях одновременно.

Исходной информацией является матрица типа "объект-признак", где объектами являются города, а признаками — показатели демографической ситуации (численность населения, общий коэффициент рождаемости, общий коэффициент смертности, общий коэффициент брачности, общий коэффициент разводимости, общий коэффициент механического движения, средний размер семьи, доля лиц трудоспособного возраста). Использование этого набора показателей обусловлено специальной статистической формой текущего учета "Социально-экономический паспорт города", разработанной Госкомитетом и Госпланом РСФСР.

При исследовании демографической ситуации в городах для агрегирования информации (получения классов-типов) используем две модели.

Первая модель, в основе которой лежит метод главных классов [1], позволяет строить многомерную типологию, исходя из предположения о том, что исходная матрица "объект-признак"

порождается суммами "вкладов" отдельных кластеров. Полученные результаты описываются параметрами, заложенными в самой модели; оценивается вклад каждого кластера в дисперсию исходных данных и на этой основе оценивается сравнительная значимость кластеров, указываются средние значения признаков для каждого кластера, а также веса признаков в каждом кластере. Число классов определяется на основе анализа соответствующей доли дисперсии.

Характерной чертой второй используемой модели является то, что совокупность городов рассматривается как развивающаяся система. При этом имеется в виду то обстоятельство, что процесс развития демографической ситуации в городах растянут во времени и пространстве таким образом, что в один и тот же момент можно наблюдать города, находящиеся на разных стадиях демографического развития. В результате получаем типологию, которая является как бы фотографией процесса перехода объектов (городов) с одного уровня демографического развития на другой. Алгоритм построения порядковой классификации [2, 3] позволяет получить классы, упорядоченные по "похожести" на соседние классы.

Эти методы были опробированы при построении многомерных типологий всей совокупности городов Западной Сибири по показателям демографической ситуации, а также для построения многомерной типологии всех городов РСФСР по показателям городской среды. Результаты представлены в ЦЭИИ Госплана РСФСР.

Л И Т Е Р А Т У Р А

1. Миркин Б.Г. Метод главных кластеров // Автоматика и телемеханика. - № 10. - М.: Наука. - 1987. - С. 131-143.
2. Миркин Б.Г., Высоцкая Н.В. и др. Шкалы упорядочения // Моделирование в экономических исследованиях. - Новосибирск: Наука. - 1978. - С. 109-119.
3. Высоцкая Н.В. Метод порядковой классификации многомерной социально-экономической информации. - Новосибирск: Б.и. - 1979. - 9с. (Препринт/ИЭ и ОПИ СО АН СССР).
4. Терещенко О.В. Демографическая ситуация в небольших городах Западной Сибири: ИЭ и ОПИ СО АН СССР. - 1987. - С. 103-119.

ПЛАНЫ КОНТРОЛЯ ПО КАЧЕСТВЕННОМУ ПРИЗНАКУ И ХАРАКТЕРИСТИКИ РЕГЛАМЕНТОВ МАТЧЕЙ

Л.Р.Гиль, Я.П.Лумельский

В настоящем сообщении в предположениях статьи [1] показывается возможность применения полиномиальных планов двухступенчатого контроля [2] в качестве приемлемой модели регламентов шахматных матчей.

Основными группами регламентов матчей на первенство мира по шахматам являются регламент А (играется фиксированное число $n = 2\ell$ партий) и регламент В (матч продолжается до k побед, а ничьи не учитываются). Для использования вероятностных методов при описании характеристик различных регламентов матчей, следуя [1], применяются полиномиальные планы контроля и делаются следующие упрощающие предположения: p_1 — вероятность выигрыша партии чемпионом, p_2 — вероятность выигрыша партии претендентом,

$q = 1 - p_1 - p_2$ — вероятность ничейного исхода партии.

С развитием шахмат наметилась тенденция возрастания вероятности ничейного исхода партии, а при $q > 0,6$ регламенты групп А и В становятся мало приемлемыми [1]. При регламенте А вероятность счета $\ell : \ell$, когда $p_1 = p_2 = p < 0,2$ велика, что ставит претендента в существенно неравные условия с чемпионом. В случае использования регламента В математическое ожидание числа сыгранных партий до k побед может достигать и сотни при больших q .

В значительной степени вышеуказанных недостатков лишены регламенты, основанные на полиномиальных планах 2 -ступенчатого контроля $\Pi^2(n_1, \dots, n_2)$, $n_i = 2\ell_i$, $i = 1, 2$, $2 \geq 2$. В этом случае первый этап матча состоит из $2\ell_1$ партий, а при ничейном счете $\ell_1 : \ell_1$ проводится следующий этап матча и т.д.; если по результатам 2 этапов счет ничейный, то чемпион сохраняет звание. Четное число партий на каждом этапе матча обеспечивает каждому участнику одинаковое число партий

l_i , в которых он играет белыми фигурами.

В случае двухступенчатого плана $\Pi^2(n_1, n_2)$ вероятность ничейного исхода и математическое ожидание числа сыгранных партий при $p_1 = p_2 = p$ находятся соответственно по формулам

$$P(p/\Pi^2) = \sum_{x=0}^{l_1} \sum_{y=x}^{x+l_2} \frac{n_1! n_2! p^{2y} (1-2p)^{n_1+n_2-2y}}{[x!]^2 [(y-x)!]^2 (n_1-2x)! (n_2-2(y-x))!} \quad (1)$$

$$E_p(\tau/\Pi^2) = n_1 + n_2 \sum_{x=0}^{l_1} \frac{n_1}{[x!]^2 (n_1-2x)!} p^{2x} (1-2p)^{n_1-2x} \quad (2)$$

Ниже в табл. I приведены для иллюстрации численные значения характеристик (1), (2) и вероятность ничейного исхода при регламенте A, которую обозначим $P(p/\Pi^1)$.

Таблица I

Характеристики P	$P(p/\Pi^1)$		$P(p/\Pi^2)$		$E_p(\tau/\Pi^2)$	
	$n = 24$	$n = 32$	$n_1 = n_2 = 12$	$n_1 = n_2 = 16$	$n_1 = n_2 = 12$	$n_1 = n_2 = 16$
0,10	0,1843	0,1591	0,0704	0,0518	15,18	19,64
0,20	0,1284	0,1113	0,0328	0,0247	14,17	18,51
0,30	0,1044	0,0906	0,0215	0,0162	13,76	18,04
0,40	0,0902	0,0783	0,0160	0,0121	13,52	17,76

Из табл. I видно, что использование двухступенчатого плана контроля $\Pi^2(n_1, n_2)$ вместо одноступенчатого Π^1 (регламент) с $n_1 = n_2$, $n = 2n_1$ позволяет сократить математическое ожидание числа играемых партий более чем в полтора раза при уменьшении вероятности ничейного исхода в несколько раз. Имеется возможность применения в качестве регламентов матчей усеченных планов двухступенчатого контроля Π_{yc}^2 , а также планов Π^3 , Π_{yc}^3 .

Л и т е р а т у р а

1. Гусев А.Л., Лумельский Я.П. Статистические оценки и вероятностные характеристики для регламентов шахматных матчей. В сб. "Многомерный статистический анализ и моделирование реальных процессов". М. Наука, 1989.
2. Лумельский Я.П. Статистические оценки результатов контроля качества. М. Изд-во стандартов, 1979.

ПРОГНОЗИРОВАНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ В РАМКАХ КЛАССИФИКАЦИОННО-ТИПОЛОГИЧЕСКИХ МОДЕЛЕЙ

Т.Р.Минина, Т.В.Маслякова, Е.В.Фример

Рассматривается задача прогнозирования для объектов, информация о состоянии которых в различные моменты времени представлена в виде матрицы "объект-признак". Предполагается, что 1) каждый из исследуемых объектов представлен лишь короткими временными рядами характеристик его состояний в исходном признаковом пространстве; 2) количество изучаемых объектов достаточно велико, а совокупность их состояний образует репрезентативную выборку из генеральной совокупности всех возможных состояний исследуемого процесса.

Методами многомерного шкалирования, реализованными в системе анализа социально-экономических данных ИСЭП АН СССР [1], строится k -мерное модельное фазовое пространство. Прогнозирование состояния изучаемого объекта проводится в этом пространстве, структура которого учитывает взаимосвязь исходных характеристик изучаемых объектов. При этом априорных предположений о форме взаимосвязи не делается. Если размерность модельного пространства не превосходит 3, то применяются методы визуализации.

При таком подходе возникают трудности, связанные с переходом от экстраполяционной точки модельного пространства к выбору прогнозной точки исходного (m -мерного) пространства. Такой переход, вообще говоря, неоднозначен. При использовании в исходном признаковом пространстве евклидовой близости прообразом каждой точки модельного пространства является некоторое квадратичное многообразие размерности $(m-k)$ [2]. Для решения указанной задачи предлагается ряд критериев, основанных на требовании максимально точной передачи структур состояний в исходном и модельном пространствах, а также максимальной гладкости траекторий состояний рассматриваемого объекта в m -мерном признаковом пространстве. В этом случае задача прогнозирования сводится к минимизации предложенных критериев при явно указанных квадратичных ограничениях.

В докладе обсуждаются разработанное математическое обеспечение, а также результаты применения предложенного подхода при прогнозировании отраслевой структуры народного хозяйства для конкретных городов и регионов областного ранга РСФСР [3]. Приводится сравнительный анализ предложенного подхода и методов независимой экстраполяции исходных характеристик исследуемого объекта.

Л и т е р а т у р а

1. Система анализа социально-экономических данных /Каныгин Г.В., Перекрест В.Т., Сивагинский С.В., Хачатурова Т.В.-Л., 1985.
2. Каныгин Г.В., Минина Т.Р., Перекрест В.Т. и др. Человеко-машинные методы анализа социально-экономических данных //Человеко-машинные системы обеспечения социально-экономических исследований.-Л.:Наука,1987.-С.92-135.
3. Крупный социалистический город: структурные аспекты развития /Отв.ред. Н.Т.Агафонов.-Л.:Наука,1987.-142 с.

МЕТОДЫ ЦЕЛЕНАПРАВЛЕННОГО ПРОЕЦИРОВАНИЯ В ВЫЯВЛЕНИИ ТИПОЛОГИИ ПОТРЕБИТЕЛЬСКОГО ПОВЕДЕНИЯ СЕМЕЙ

А.А. Митоян

В условиях совершенствования хозяйственного механизма, и в частности, при переходе к экономическим методам управления народным хозяйством, к предоставлению все большей финансовой и административной самостоятельности предприятиям, к организации социалистического рынка продуктов резко возрастает роль методов и моделей, позволяющих проводить адекватный анализ и прогноз структуры потребительского спроса, динамики объемов продаж отдельных групп товаров и услуг.

Предлагаемый нами подход к решению этой актуальнейшей проблемы основан на синтезе двух методологических приемов: (1) приема, основанного на многомерном статистическом анализе семейных бюджетов и несколько развитом нами методе выявления основных типов потребительского поведения с одновременным определением главных типобразующих признаков [1] и (2) приема, основанного на интервальном регрессионном прогнозе объемов продаж отдельных благ и услуг по данным торговой статистики [2].

Развитие подхода [1] состоит, во-первых, в использовании современных методов целенаправленного проектирования [3] в целях формирования предварительных рабочих гипотез о числе и характере основных типов потребителей и потребительского поведения, и во-вторых, в учете динамики цен на агрегированные группы продуктов и услуг и в стыковке методов (1) и (2) со схемой дифференцированного баланса доходов и потребления населения [4].

Модельные и экспериментальные расчеты подтвердили, что поиск плоскости проецирования, нацеленный на оптимизацию критериев, специфицированных в соответствии с условиями максимального сохранения (после проецирования исходных многомерных данных) именно кластерной структуры анализируемого массива [3] с.492, дает гораздо более точные и выразительные результаты, чем ранее использовавшееся проецирование на плоскость первых двух главных компонент.

Частичная экспериментальная апробация предложенного подхода осуществлена на данных семейных бюджетов и торговой статистики ГОСКОМСТАТА Арм ССР за 1971, 1981 и 1986 годы с использованием имевшегося в ЦЭМИ АН СССР и специально дополнительно разработанного нами программно-алгоритмического обеспечения и вычислительной техники ЕС-1060, WANG-2200 и IBM PC/AT.

ЛИТЕРАТУРА

- 1 Типология потребления. Под ред. С.А.Айвазяна и Н.М.Римашевской. М.: Наука, 1978. - 175 с.
- 2 А.А.Митоян. Интервальное прогнозирование в условиях неопределенности распределения регрессионных остатков и неоднородности выборок. В сб.: Применение многомерного статистического анализа в экономике и оценке качества продукции. Тарту, изд-во АН ЭССР, 1985, с. 181.
- 3 С.А.Айвазян, В.М.Бухштабер, И.С.Енюков, Л.Д.Мешалкин. Прикладная статистика: классификация и снижение размерности. М.: Финансы и статистика, 1989. - 608 с.
- 4 Карапетян А.Х. Доходы и потребление населения СССР. М.: Статистика, 1980. - 272 с.

ПРИМЕНЕНИЕ МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА К ПРОГНОЗИРОВАНИЮ РАЗВИТИЯ И К ИССЛЕДОВАНИЮ СТАБИЛЬНОСТИ МНОГОПАРТИЙНЫХ ПОЛИТИЧЕСКИХ СИСТЕМ

И.Ф.Петерсен

Пусть личные взгляды, стремления и т.п. каждого индивида (избирателя) общества, по отношению к стратегии развития данного общества (к соотв. нормам, законам, планам и пр.), как и каждой из k функционирующих в данном обществе партий (их платформы, стратегии) достаточно полно характеризуются точками в R^n . Всё разнообразие политического сознания и стремлений общества определяется тогда соответствующим вероятностным распределением $G(y)$ на R^n , а образующие политическую систему общества партии изображаются k точками x_1, \dots, x_k в этом же пространстве. Пусть еще известна функция потери $f(x, y)$, $f: R^n \times R^n \rightarrow R^+$, выражающая степень неудовлетворенности, несогласия индивида y со стратегией x развития общества. При отсутствии более подробной модели можно положить, например, $f(x, y) = \|x - y\|^2$. Если $Y_i (i=1, \dots, k)$ — подмножество индивидов, которым суждено пользоваться стратегией x_i , то суммарная неудовлетворенность общества системой партий выражается функцией

$$F(\bar{x}, \bar{Y}) = \sum_{i=1}^k \int_{Y_i} f(x_i, y) dG(y),$$

где $\bar{x} = (x_1, \dots, x_k)$, а $\bar{Y} = \{Y_1, \dots, Y_k\}$ — разбиение R^n . Обозначим через $Y_i(\bar{x})$ подмножество таких индивидов, которые среди x_1, \dots, x_k предпочитают стратегию x_i , т.е. $Y_i(\bar{x}) \subset \{y/y \in R^n, f(x_i, y) \leq f(x_j, y)\}$, и через $\bar{Y}(\bar{x})$ — соответствующее разбиение Y . Если каждый индивид общества имеет свободу выбирать среди функционирующих партий, какую поддержать, и если каждая партия приведет

свою платформу в согласие с поддерживающим ее множеством индивидов $Y_1(\tilde{x})$ в смысле минимизации потерь, то вектор, описывающий платформы партий \tilde{x} стремится к равновесному состоянию \tilde{x}^* , характеризующемуся условиями

$$F(\tilde{x}^*, \tilde{Y}(\tilde{x}^*)) \leq F(\tilde{x}^*, \tilde{Y}) \quad \text{при любом разбиении } Y,$$

$$F(\tilde{x}^*, \tilde{Y}(\tilde{x}^*)) \leq F(\tilde{z}, \tilde{Y}(\tilde{x}^*)) \quad \text{при любых платформах } z_1, \dots, z_k.$$

Равновесные состояния являются стационарными точками функционала $F(\tilde{x}, \tilde{Y}(\tilde{x}))$, а их устойчивость заключается в положительной определенности блочной матрицы

$$\left(\frac{\partial^2}{\partial x_i \partial x_j} F(\tilde{x}^*, \tilde{Y}(\tilde{x}^*)) \right).$$

При плавном изменении распределения G равновесное состояние может потерять стабильность, что приводит к катастрофе в смысле Тома политической системы общества.

В свете изложенного статистическое исследование направлений развития платформ партий и стабильности политической системы сводится к выборочному оцениванию распределения G и платформ партий в данное время и к построению на их основе оценок изменений платформ $\hat{z}_i - \hat{x}_i$, где \hat{z}_i вектор, минимизирующий в среднем потери множества $Y_1(\hat{x})$. Далее проверяется гипотеза, что эти изменения равны нулю, т.е. что политическая система находится в равновесном состоянии, и при принятии этой гипотезы проверяется еще вторая гипотеза о стабильности системы партии. При оценивании G может быть применен метод многомерного шкалирования или, при подходящей априорной параметризации — метод главных компонент.

В качестве иллюстрации рассматриваются данные о голландском парламенте за 1972 г. Обсуждаются перспективы проведения такого статистического анализа в Эстонии.

Л и т е р а т у р а

1. Thom R. Structural Stability and Morphogenesis.- Benjamin-Addison Wesley, New York, 1975.
2. Петерсен И. Куда ведет демократия?.- В кн. Численные методы и оптимизация. Изд. "Валгус", Таллинн, 1988, с.152-157.

СТАТИСТИЧЕСКИЕ ПОТЕНЦИАЛЫ

О.В. Староверов

Статистические или жизненные потенциалы поколения появились для выяснения вопроса о средней продолжительности предстоящей жизни людей, принадлежащих данному поколению [1]. Затем рассматривались, например, в работах А.С.Первушина трудовые потенциалы поколения, т.е. средние продолжительности активной (трудовой) жизни людей. Решение всех этих задач обслуживалось чисто демографическими моделями, в которые включались коэффициенты трудовой активности.

Однако многие важные вопросы социальной и экономической жизни остаются без ответа. Например, вопрос о среднем стаже работника, начавшего свою трудовую жизнь шахтером до его выхода на пенсию. Вопрос о доле людей, занятых в разных сферах деятельности данного региона и прекративших работу из-за выезда из региона, выхода на пенсию или смерти, также требует ответа. Но для этого требуются общие модели, отличные от демографических.

Движение населения в наиболее общем виде достаточно качественно отражают марковские ветвящиеся процессы с несколькими типами частиц и непрерывным временем [2].

В сообщении имеется решение поставленных вопросов для марковских моделей с непрерывным временем. В качестве числовой иллюстрации приведены данные о средних сроках пребывания в регионе в зависимости от группы (например, отрасли) до выезда из него, выхода на пенсию или смерти и соответственно о долях лиц, принадлежащих к разным группам и выехавших из региона, вышедших на пенсию или умерших до выхода на пенсию.

Л и т е р а т у р а

1. Фильрозе Э. Очерки потенциальной демографии. М. 1975.
2. Староверов О.В. Модели движения населения. "Наука", М. 1979.

СРЕДСТВА ТИПОЛОГИЧЕСКОГО МОДЕЛИРОВАНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ ИНФОРМАЦИИ

Т.В.Хачатурова, Е.М.Ильин, В.Т.Перекрест

При моделировании социально-экономических процессов методами многомерного шкалирования функция K -мерного шкалирования $f: X \rightarrow R^K$ может быть построена как Z_K -отображение для пространства исходных объектов $X = (X, z, \mu)$ с функцией близости z и вероятностной мерой μ [1]. Построенное при этом модельное фазовое пространство $H = f(X) \subset R^K$ реализует систему взаимосвязей между исходными характеристиками изучаемых объектов [2]. В нем удобно описываются (в ряде случаев - визуализируются) отдельные объекты и классы объектов исходного пространства, траектории их движения, эталонные ("идеальные") объекты, а также строится экстраполяция коротких временных рядов.

Обсуждаются программные средства, поддерживающие анализ данных с использованием фазового пространства H : построение начального приближения для f , решение оптимизационной задачи Z_K -шкалирования, кластеризация объектов в X , построение прогноза. Рассматривается комплекс программ визуализации функциональной модели шкалирования (ВИМШ). Комплекс ВИМШ, 1-я версия которого реализована для IBM PC XT, обеспечивает следующие возможности:

- прием задания от пользователя в режиме графического диалога;
- выбор заказанных объектов и признаков, вычисление (для выбранных объектов) функции шкалирования и ее производных по исходным признакам;
- вывод на экран заказанных элементов изображения, в том числе: одновременное отображение всех объектов изучаемой выборки на шкалу, с целью визуального определения сгущений и крайних точек; отображение на шкалу всех объектов того или иного класса; отображение на шкалу одного объекта-представителя для каждого класса; помещение на шкалу точек, отображающих положение одного и того же объекта в различные

моменты времени (с целью изучения возможных траекторий развития объектов); отображение на шкалу того или иного идеального объекта и его сопоставление с образами реальных объектов; отображение на шкалу реального объекта при различных смещениях его в том или ином направлении в исходном пространстве; вывод (в дополнительном окне) сопутствующих числовых характеристик классов объектов или отдельных объектов.

Комплекс ВИМШ работает с базой данных — объектами, описанными через признаки исходного пространства X . В эту же базу помещается информация о результатах кластеризации объектов в исходном пространстве, а также сведения об изменении положения объектов в различные моменты времени. В I-й версии ВИМШ для этих целей используется СУБД αBASE III или РЕБУС.

Параметры функционального отображения шкалирования и результаты кластеризации объектов в исходном пространстве определяются в настоящее время средствами [3] на центральной ЭВМ, а затем передаются на IBM PC по каналу связи для изучения средствами ВИМШ. Разрабатываются программные средства построения параметров функциональной модели шкалирования непосредственно на IBM PC.

Л и т е р а т у р а

1. Перекрест В.Т. Нелинейный типологический анализ социально-экономической информации.—Л.:Наука,1983.—176 с.
2. Крупный социалистический город: структурные аспекты развития /Отв. ред. Н.Т.Агафонов.—Л.:Наука,1987.—142 с.
3. Система анализа социально-экономических данных /Каныгин Г.В., Перекрест В.Т., Сивашинский С.В., Хачатурова Т.В.—Л.,ИСЭП АН СССР,1985.—48 с.

АВТОМАТИЗИРОВАННОЕ РАБОЧЕЕ МЕСТО
ЭКОЛОГА-ИССЛЕДОВАТЕЛЯ

В.Н.Виноградова, А.Г.Закиров, Л.Л.Фрлова

В связи с расширением работ по экологическим исследованиям, появлением различных служб охраны природы на уровне как государственных, так и общественных организаций, становится актуальной проблема обеспечения таких исследователей средствами хранения и обработки данных.

Этапы хранения, первичной обработки, определения наличия зависимостей между различными показателями природной среды и интенсивностью антропогенного воздействия на природу и оценки репрезентативности данных являются базовыми. Они являются необходимыми при любой направленности использования исходного материала: при оценке и прогнозировании состояния природной среды, определении материального ущерба от нарушений экосистемы, создании математических моделей экологических процессов, нормирование и т.д. Массовое использование методов первичной обработки информации требует наличия удобных программных средств.

Одним из таких средств для выполнения функций хранения и математической обработки результатов является симбиоз системы управления базой данных и пакета программ по статистическому анализу. При этом используемая СУБД должна быть достаточно широко распространена или легко доступна, а пакет статистических программ (ПСП) должен обладать достаточной полнотой используемых методов.

Из условия совместимости СУБД и ПСП следует, что СУБД должна быть реляционного типа, т.к. с одной стороны табличное представление данных удобно для пользователей, с другой стороны статистическая обработка также привязана к табличному представлению данных.

В Казанском государственном университете разработан пакет программ СТАТ, объединяющий в себе функции реляционной СУБД СМТ с языком запросов QBE [1] и пакета прикладных программ "Статистическая обработка данных" [2].

Перед выполнением статистического анализа экологу-исследователю предоставляется возможность выборки индивидов по условию,

допустимому в СУБД СМТ, выборки индивидов, не имеющих нулевых значений признаков, выборки индивидов, не имеющих пустых значений признаков.

Результат выборки по любому из перечисленных условий подвергается статистическому анализу.

Для проведения статистического анализа в пакете программ СТАТ реализованы следующие функции:

- Первичная статистическая обработка данных, заключающаяся в получении основных одномерных и двумерных статистик.

- Анализ главных компонент, состоящий в выявлении линейных структурных зависимостей между признаками и с выдачей графиков зависимостей на графический дисплей.

- Регрессионный анализ.

- Дискриминантный анализ.

- Дисперсионный анализ.

Простота работы пользователя в пакете программ СТАТ, которая достигается с помощью диалога, построенного на использовании программируемых функциональных клавиш для выполнения выборки и запоминания данных, управления работой системы, позволяет использовать его при обучении студентов экологов.

Литература

1. Дейт К. Введение в системы баз данных. - М.
2. Тез.докл. II Всесоюзной научно-технической конференции "Применение статистического анализа в экономике и оценке качества продукции". - Тарту, 1985. - Ч.1. - 255 с.; Ч.2. - 230 с.

СТАТИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ КОМПОНЕНТ ЗАГРЯЗНЕНИЯ АТМОСФЕРЫ

Е.А.Зеленюк

Для изучения процессов загрязнения атмосферы необходимы оценки уровней загрязненности, характерных для исследуемого района, т.е. оценки фоновых уровней загрязненности. Получение таких оценок затрудняется влиянием на уровни загрязненности большого числа физических, метеорологических и географических факторов. В предлагаемой работе извлечение информации о процессах загрязнения из рядов наблюдений связывается с разработкой и использованием статистических представлений о данных, а именно как задача получения устойчивых статистических характеристик /1,2/.

Анализ данных станций фоновых мониторинга, расположенных в различных регионах (Швеция, Белоруссия, Средняя Азия, Казахстан), для различных типов загрязняющих веществ (диоксиды серы, свинец, пыль) показал, что графики функций распределения на нормальной вероятностной бумаге, с удаленным линейным трендом, имеют характерную форму "сгустков", характеризующих отклонения реального распределения от логнормальности, причем каждый из сгустков может рассматриваться как описание части многомодального распределения /1-4/. Была построена математическая модель такого распределения и описаны его свойства.

Получаемая из рядов наблюдений информация представляется в виде "интервалов группирования" концентраций, характеризующих различные компоненты загрязнения атмосферы. Разработанный аппарат сравнения интервалов и выделения на этой основе устойчивых компонент загрязнения позволил сделать ряд выводов о процессах формирования фоновых загрязнений, и провести сравнение фоновых уровней загрязненности в континентальных масштабах.

Разработанная методика в течении 1987-88 г.г. использовалась при анализе аэрозольной загрязненности в районе Чернобыльской АЭС. Решалась задача восстановления обобщенной картины загрязненности в локальных масштабах (15х15 км).

Был получен ряд новых результатов относительно связи компонент загрязнения с метеоусловиями, источниками загрязнения, т.п. Сеть пробоборота представлена 27 датчиками, 23 из которых расположены на расстоянии до 7,5 км от венттрубы второй очереди ЧАЭС. Показано, что резкие колебания активности аэрозолей на расстояниях до 2 км связаны с характером израспространения, и задача сводится к изучению поведения "пылящих пятен" и нахождению таких статистик, которые позволили бы построить обобщенную картину загрязненности территории. Проведенный анализ интервалов группирования активностей позволил установить направления поступления аэрозолей различных активностей в точки пробоборота, на основании чего были построены карты среднесезонной загрязненности ближней зоны ЧАЭС – антропогенной, связанной с проведением работ (уровни до 5.000 ед) и "Фоновой", определяемой характером послеаварийной загрязненности территории (уровни до 5 ед.)

Л и т е р а т у р а .

1. Ю.А.Израэль, Ф.Я.Ровинский, М.Я.Антоновский, В.М.Бухштабер, Е.А.Зеленюк, Ю.П.Черханов, К статистическому обоснованию компонента загрязнения атмосферы в фоновом районе, ДАН, 1984, т.276, №2, стр.334-337.
2. М.Я.Антоновский, В.М.Бухштабер, Е.А.Зеленюк, Анализ фонового загрязнения атмосферы на основе многомодальных распределений, Труды III Международного симпозиума "Комплексный глобальный мониторинг...", Л., Гидрометеиздат, 1986.
3. Ю.А.Израэль, М.Я.Антоновский, В.М.Бухштабер, Е.А.Зеленюк, К проблеме выявления и оценки фоновых уровней загрязнения атмосферы, ДАН, 1987, т.292, №2, стр.309-314.
4. М.Я.Антоновский, В.М.Бухштабер, Е.А.Зеленюк, A statistical model of background air pollution frequency distributions, IIASA Working Paper, Austria, Laxenburg, 1988.

ПРИМЕНЕНИЕ УНИВЕРСАЛЬНОГО ПО ПС ДЛЯ ПРОБЛЕМНО-ОРИЕНТИРОВАННОГО СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ О СОСТОЯНИИ ОКРУЖАЮЩЕЙ СРЕДЫ

А.А.Зубенко, Г.М.Зубенко

Применение методов статистического анализа к данным о среде обусловлено сложностью, многообразием и вероятностной природой процессов, происходящих в окружающей среде. Необходимость использования методов прикладной статистики при анализе климатических данных следует из статистического определения климата (как среднего состояния атмосферы, океана, льда и земной поверхности в определенном регионе за определенный период времени). Предсказываемые численными моделями серьезные экологические последствия от изменения содержания примесей в атмосфере привели к необходимости проведения всестороннего анализа данных мониторинга среды с использованием всего арсенала методов прикладной статистики и имеющихся данных. Для включения опыта и знаний геофизика в процесс анализа данных с использованием современных достижений прикладной статистики необходимо создать множество интерактивных проблемно-ориентированных систем анализа на базе существующих пакетов статистических программ. Для этого были разработаны средства взаимодействия подсистем хранения и анализа данных в рамках центра данных о среде, а также средства взаимодействия с пользователем, обеспечивающие проведение такого анализа силами геофизиков.

Использование специфики определенной области знания может повысить эффективность информационной системы, т.к. лучшая форма организации данных зависит от содержания последних и от требуемого вида доступа к информации. Так как статистический анализ (не требующий модификации самих данных) – основной предполагаемый вид обработки, а добавления

новых порций данных достаточно редки, были выбраны поисковые структуры на основе частично-инвертированных файлов (указанный способ хранения данных реализован, например, в СУБД СПЕКТР или ДИСОД). Такой способ хранения информации обеспечивает быстрый поиск и выборку данных за счет относительно большого времени выполнения операций добавления и модификации. Для оперативной выборки требуемых порций информации из базы данных программами статистического анализа был разработан интерфейс системы SAS с СУБД СПЕКТР на основе языка запросов NATURAL в виде отдельной процедуры системы SAS. Чтобы облегчить программисту создание проблемно-ориентированных подсистем, использовалась единая концепция управления вычислительным процессом и средств общения с пользователем: единая среда – система SAS – позволяет унифицировать обращение к модулям различных пакетов и представление массивов данных, единое средство взаимодействия с пользователем – генератор прикладных программ ДИАГЕН [1] – обеспечивает ведение диалога на языке предметной области и накопление сценариев анализа, оставляя за кадром генерацию заданий на языках ППП и вызов требуемых программ.

Использование системы SAS для хранения самих данных в среде было отвергнуто, т.к. управление данными системы SAS не обеспечивает компактности хранения данных, одновременную работу нескольких пользователей с одной базой данных, не поддерживает логически связанных модификаций и процедур восстановления (в рамках центра данных производится периодическое пополнение информации одновременно с обработкой).

Л и т е р а т у р а

1. Бухштабер В.М., Зеленик Е.А., Зубенко А.А. Конструирование интерактивных систем анализа данных. М.: Финансы и статистика, 1989.

МНОЖЕСТВЕННЫЙ ЛИНЕЙНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ ВЛИЯНИЯ ХИМИЧЕСКИХ КЛАССОВ ПЕСТИЦИДОВ НА ЗДОРОВЬЕ НАСЕЛЕНИЯ

А.Б.Качинский, В.И.Польченко, Н.П.Сарническая

В данной работе рассмотрено приложение множественной линейной регрессии для оценки состояния здоровья населения, где в качестве зависимой переменной рассматривалось 32 нозологические формы заболеваний детей возрастом от 0 до 14 лет, а независимыми переменными служили 6 химических классов пестицидов (фосфорорганические, хлорорганические, карбоматы, медьсодержащие, гербициды и прочие) для 11 различных регионов страны. Предполагалось, что все переменные суть случайные величины, имеющие совместное многомерное нормальное распределение и эффекты влияния каждого химического класса на заболеваемость складываются, тогда нашу модель можно записать в виде множественной линейной регрессии

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_6 x_{i6} + e_i, \quad (1)$$

где β_0, \dots, β_6 - неизвестные параметры, а e_i - независимая случайная ошибка, распределенная по закону $N(0, \sigma^2)$.

С помощью общей линейной модели мы решили ряд задач, которые перечислены в таблице I, с учетом следующих предположений относительно случайных компонент e_i :

А. Случайные величины e_i независимы, имеют средние значения, равные нулю, и постоянную дисперсию σ^2 , которая не зависит ни от β_i , ни от X_{ji} .

В. Случайные величины e_i независимы и нормально распределены со средними, равными нулю, и с постоянной дисперсией σ^2 , которая не зависит ни от β_i , ни от X_{ji} .

Результаты моделирования показали, что для таких регионов страны как Узбекистан, Туркмения, Краснодарский край отмечается высокий коэффициент множественной корреляции (0,7 - 0,9), что указывает на сильную линейную зависимость между практически

ки всеми формами заболеваемости и используемыми химическими классами пестицидов.

Проверка гипотезы $H_0: \beta_1 = \beta_2 = \dots = \beta_6 = 0$, которую мы рассматривали как гипотезу о том, что независимые переменные $\{X_1, X_2, \dots, X_6\}$ не улучшают предсказания Y относительно $\hat{y} = \bar{y}$, дала возможность построить уравнения регрессии (1) для таких заболеваний как пневмония, отит хронический, туберкулез органов дыхания, холецистит, шизофрения и др.

Гипотеза $H_0: \beta_k = 0, k = 1, 6$ рассматривалась нами как гипотеза о том, что переменная X_k не улучшает предсказание Y сравнительно с предсказанием, полученным по $(p-1)$ относительным переменным. Показано, что пневмония у детей до I года определяются, в основном, фосфорорганическими пестицидами и карбоматами; отит хронический - фосфорорганическими пестицидами; туберкулез органов дыхания - медь-содержащими пестицидами.

Таблица I

№ : п/п :	Задачи	: Предположения о : случайных вели- : чинах
1.	Вычисление точечных оценок $\hat{\beta}_i$ для β_i	A или B
2.	Вычисление точечной оценки $\hat{\sigma}^2$ для σ^2	A или B
3.	Построение доверительных интервалов для μ_i	B
4.	Построение доверительных интервалов для μ_y	B
5.	Проверка гипотезы $H_0: \beta_1 = \beta_2 = \dots = \beta_6 = 0$	B
6.	Проверка гипотезы $H_0: \beta_k = 0, k = 1, 6$	B

Приведенные выше данные носят предварительный характер, которые будут уточняться по мере поступления новых материалов.

АНАЛИЗ ЭКСПЕРТНЫХ МНЕНИЙ О ПРИЧИНАХ
И МЕРАХ ПРЕДОХРАНЕНИЯ ЧУДСКОГО ОЗЕРА
ОТ ЗАГРЯЗНЕНИЯ

А.Маамяги

В рамках темы "Имитационное моделирование динамики экосистем мелководных водоемов под влиянием хозяйственной деятельности в регионе" были опрошены эксперты - научные работники и исследователи, занимающиеся изучением состояния Чудского озера, о проведении наиболее важных по их мнению мероприятий (с точки зрения сохранения экологического равновесия озера). Анкета была составлена Х.Симм.

Анкета содержала перечень 19 мероприятий (очистные сооружения для города Тарту, оптимизация удобрительного комплекса, уменьшение применения пестицидов итд.). Была предусмотрена возможность уточнять вопросы, так и возможность дополнить их список. Также было нужно оценить степень необратимости происшедших изменений.

Ответы требовалось представить в виде упорядоченных разбиений представленных 19 вопросов не более, чем на четыре класса (первый класс - первостепенные мероприятия, второстепенные - второй класс, менее важные - третий, не имеющие значения - четвертый).

Полученные ответы анализировались как с точки зрения компетентности ответивших экспертов, так и с точки зрения существования "истинного упорядоченного разбиения" рассматриваемых мероприятий [1]. Ответы были разделены случайным образом на две совокупности и анализировались отдельно.

"Истинное разбиение" содержало три класса.

Л и т е р а т у р а

1. Маамяги А.В. Некоторые задачи статистического анализа классификаций. - Таллинн, 1982. - 23 с. (Препринт АН ЭССР).

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ
ДАННЫХ ФОСФОРИТНЫХ МЕСТОРОЖДЕНИЙ
ТООЛСЕ И КАБАЛА

А. Маамяги, С. Малыгин

Пусть даны выборочные значения случайных чисел $\{X_i\}_{i=1}^{10}$. Требуется найти вероятность событий $P(X_1 \geq a_1, X_4 \leq a_4, X_5 \leq a_5)$ и $P(X_6 \geq a_6, X_9 \leq a_9, X_{10} \leq a_{10})$ ($P(X_1 \geq a_1, X_2 \leq a_2, X_5 \leq a_5)$, $P(X_6 \geq a_6, X_7 \leq a_7, X_{10} \leq a_{10})$), $\{a_i\}_{i=1}^{10}$ - фиксированные числа.

X_1 - содержание P_2O_5 в руде Тоолсе.

X_6 - содержание P_2O_5 в руде Кабала-Западный.

X_2 - содержание окси магния в руде Тоолсе.

X_7 - содержание окси магния в руде Кабала-Западный.

X_3 - содержание окси железа в руде Тоолсе.

X_8 - содержание окси железа в руде Кабала-Западный.

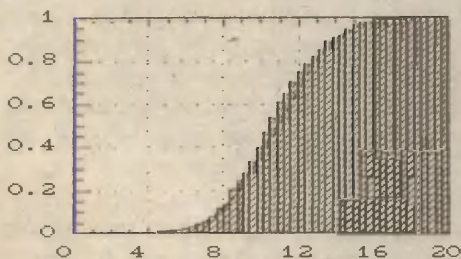
$X_4 = X_2/X_1$, $X_5 = X_3/X_1$, $X_9 = X_7/X_6$, $X_{10} = X_8/X_6$. Искомые оценки

$P(X_1 \geq a_1, X_4 \leq a_4, X_5 \leq a_5)$ и $P(X_6 \geq a_6, X_9 \leq a_9, X_{10} \leq a_{10})$

определяют долю пригодного к химической обработке сырья, являясь оценкой качества руды. Выборочные данные предоставлены Р. Пязоком.

Определены совместные распределения (X_1, X_4, X_5) и (X_6, X_9, X_{10}) .

$(X_1, X_2, X_5), (X_6, X_7, X_{10})$.



Теоретическая и эмпирическая функция распределения X_1 .

ОЦЕНИВАНИЕ ГЛОБАЛЬНЫХ СТАТИСТИЧЕСКИХ ПАРАМЕТРОВ МОЛЕКУЛЯРНОЙ ЭВОЛЮЦИИ

Т. Мелс

Молекулярные филогенетические деревья строят по матрицу расстояний между биологическими видами. Для сравнения видов используются различия в химической структуре белков, нуклеиновых кислот и др. веществ. Расстоянием служит число несопадающихся элементов в цепных молекулах этих веществ. Считают, что любые сравниваемые молекулы развивались в ходе эволюции из общего предка путем последовательных случайных возмущений. Каждое возмущение состоит из случайного выбора места (локуса) в молекуле и из случайной замены элемента в этом локусе (результатирующих изменений ожидается меньше, чем возмущений). Поток возмущений предполагается стационарным случайным процессом с независимыми приращениями.

Пусть число возмущений за период t имеет отрицательное биномиальное распределение или распределение Неймана типа A с ожиданием $E\chi = (\lambda t)\kappa$ и дисперсией $D\chi = (\lambda t)\kappa(\kappa + 1)$, где λt и κ — параметры распределения (локальный и глобальный соответственно). Параметр κ предполагается постоянным, а если λ постоянный и эволюция протекала в реальном времени, то филогенетическое дерево называется регулярным.

Для эволюционной теории большой интерес представляет глобальный параметр. Предлагаем метод оценивания κ в случае, где истинное дерево регулярное и ее топология задана. Схема оценивания: (1) преобразовать имеющиеся расстояния между современными видами в соответствующие числа возмущений; (2) по этим числам оценить обобщенным линейным методом наименьших квадратов числа возмущений x_i на дугах дерева, сначала без предположения регулярности (получим оценки \tilde{x}_i), а затем предполагая регулярность (получим \bar{x}_i); (3) используя статистику $\langle \sum (\tilde{x}_i - \bar{x}_i)^2, x_1, \dots, x_k \rangle$ (k — число дуг дерева) можно вычислить оценку для κ . Конечные но громоздкие формулы получены модифицированным методом моментов. Статистические свойства предлагаемого метода изучены путем статистического моделирования.

ПРИНЦИПЫ ОРГАНИЗАЦИИ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ МЕДИКО-ГЕНЕТИЧЕСКИХ ДАННЫХ СРЕДСТВАМИ ПРОГРАММНОЙ СИСТЕМЫ О М Е Г А

О.И.Рябова, И.Б.Марченкова

Прикладная Программная Система Общего и Медицинского Генетического Анализа (ППС ОМЕГА) /1/, реализованная на языках ПЛ/I и АССЕМБЛЕР в среде СВМ ЕС ЭВМ, предназначена для автоматизации анализа данных медицинских профилактических исследований. Такие исследования, основанные на структурном генетико-экологическом подходе к решению задач прогноза $1/2$ /, включают в себя этап выявления патогенетических форм предрасположенности к полигенным заболеваниям. ППС ОМЕГА позволяет осуществить решение возникающих на данном этапе генетико-статистических задач путем обработки семейных данных. Система признаков (не более 1000) может включать признаки эко- и фенотипического происхождения, представленные в номинальной, порядковой и количественной шкалах.

Стратегия работы пользователя с системой определяется методическими и техническими средствами организации процесса решения задач. Методическими средствами, доступными пользователю, является набор типовых задач, отражающий ключевые моменты процесса выявления патогенетических форм как этапы решения задачи "Генетико-экологический анализ структуры популяции"(1). Типовыми задачами первого уровня являются: "Поиск маркеров"(2) и "Построение фено- и экокласов"(3). В рамках решения задачи (2) осуществляется изучение свойств и структуры пространства признаков, используемых в задаче (1). Решение задачи (3) позволяет выявить фено- и экотипическую структуру популяции, связанную с исследуемым заболеванием. Процесс решения задачи (2) и (3) формируется на основе решения трех типовых задач второго уровня: "Выявление свойств

признаков"(4), "Построение пространства признаков"(5), "Построение классов"(6). Решение любой из задач (4)–(6) связано с определением следующих свойств признаков: независимости, детерминированности, интегрированности, информативности первого и второго порядка. Процесс выявления этих свойств может рассматриваться как решение некоторой типовой задачи третьего уровня в структуре вычислительного процесса. Смысл каждого свойства задан реализованным в ППС ОМЕГА генетико-статистическим методом. Оценка независимости признаков основана на определении фенотипических, генетических и средовых связей пар признаков. Степень генетической и средовой детерминированности признака устанавливается путем использования методов генетико-дисперсионного анализа. Смысл свойства интегрированности определяется реализованным способом построения производных признаков (индексы типа отношений). Поиск признаков, обладающих свойством информативности первого порядка, предполагает оценку фенотипических и (или) генотипических ассоциаций. Информативность второго порядка – способность отдельного признака влиять на эффективность классификаций.

Решение пользовательской задачи осуществляется в рамках следующих организационно-технических средств: диалоговые языки управления ("приказов" и "распоряжений"), инструкции, описание работы системы, поддержка решаемых задач (ведение протокола, дневника, "грифельной доски").

В настоящее время ведется разработка второй версии ППС ОМЕГА в среде ПЭВМ, результат которой позволит проводить генетико-профилактические исследования более широкому кругу пользователей.

Л И Т Е Р А Т У Р А

1. Rostorzev V.N., Petrov A.A., Ryabkova O.I. *Structure and Functions of Applied Program System "OMEGA" in Medical Decision Making Diagnostic Strategies Expert Systems.* North-Holland, Amsterdam, New-York, 1985, p. 316 - 318

2. Ростовцев В.Н. Генетика и диагноз. Минск, Университетское, 1986 – 189 .

ПЛАНЫ БЛУЖДАНИЙ ПОЛИНОМИАЛЬНОГО ТИПА В МОДЕЛЯХ НАРУШЕНИЯ РИТМА СЕРДЦА

П.Н.Сапожников

Рассматриваемая математическая модель была построена при решении задачи, предложенной кафедрой терапии ФУВ ПГМИ, оценивания зависимости частоты сердечных сокращений ЧСС от ряда факторов. Информация о ЧСС каждого человека, содержащаяся в ЭКГ была предварительно редуцирована к вектору, координатами которого являлись суммарные количества сокращений: замедленной частоты (явление брадикардии), нормальной частоты и ускоренной частоты (явление тахикардии) за определённый период времени t . Кардиологи различают несколько стадий тахикардии, возможны также явления экстрасистолы и др.. Здесь рассматривается упрощенная модель, в которой все типы нарушений ЧСС сведены к двум, однако, увеличение размерности вектора нарушений не требует принципиально новых построений.

Пусть $\alpha_1 > \alpha_2 > \alpha_3$ средние длительности кардио-циклов, а $v(\tau) = (v_1(\tau), v_2(\tau), v_3(\tau))$ - вектор количеств полных циклов указанных трёх типов ЧСС за время τ . Поставим в соответствие процессу $v(\tau)$, $0 \leq \tau \leq t$, план марковских случайных блужданий по целочисленной решётке 1-го октанта R^3 с границей остановки

$$\Gamma_t = \{(n_1, n_2, n_3) : \alpha_1 n_1 + \alpha_2 n_2 + \alpha_3 n_3 \leq t\}$$

и параметром $\theta = \{\theta_{i,j}, \theta_j; i, j = 0, 1, \dots, \lceil t/\alpha_3 \rceil - 1\}$,

$$\theta_{i,j} = P\{e_1 + i e_2 + j e_3 | i e_2 + j e_3\}, \theta_j = P\{e_2 + j e_3 | j e_3\},$$

где e_i ($i=1, 2, 3$) - единичный орт i -ой координатной оси. Область изменения параметра θ с учётом естественных ограничений, будем обозначать через Θ .

Тот факт, что вся информация о поведении процесса $v(\tau)$, $0 \leq \tau \leq t$, представлена лишь его финальным значением $v(t)$, побуждает предъявить к плану блужданий следующее ограничение - граница остановки Γ_t является достаточной статистикой для

параметра $\theta \in \Theta$. Обширный класс планов марковских случайных блужданий, удовлетворяющих указанному ограничению, образуют планы блужданий, у которых вероятности переходов $S \rightarrow S + e_i + e_j$ ($i \neq j$) и $S \rightarrow S + e_1 + e_2 + e_3$ для каждой проходимой точки S не зависят от выбора пути. Такие планы будем называть планами ПТ (полиномиального типа). Вероятность перехода по фиксированной траектории из начала координат в достижимую точку $S(x, y, z)$ плана ПТ фиксированного объема n выражается через параметры $(\theta_j, \theta_{i,j}; i, j = 0, 1, \dots, n-1)$ посредством алгоритма $\Psi_{x,y,z}(\theta)$, содержащего конечное число арифметических операций. Этим алгоритмом определяются вероятности переходов в случае плана ПТ объема n с произвольной границей.

Планы ПТ охватывают класс планов Пойя и обладают свойствами аналогичными свойствам полиномиальных планов. Всякий план ПТ замкнут или нет одновременно с полиномиальным планом имеющим ту же границу остановки. Для любого параметрического множества Θ несмещёнными оценками вероятностей перехода из начала координат в проходимые точки служат траекторные оценки. Наконец, при естественных ограничениях на Θ запас параметрических функций, допускающих оптимальные несмещённые оценки на основе плана ПТ, определяется границей остановки.

Для завершения построения модели, отметим, что процесс $v(\tau)$, $0 \leq \tau \leq t$, не является марковским, поэтому соответствующий ему план остановки не может быть планом 1-го вхождения. Действительно, точки границы, удовлетворяющие неравенству $a_1 n_1 + a_2 n_2 + a_3 n_3 < t - a_3$, проходимы. Несмотря на это, этот план обладает указанными выше свойствами плана ПТ 1-го вхождения.

Применение модели опирается на предположение, что именно в параметрах (совокупности вероятностей перехода к замедленному ритму первый раз) находит отражение влияние внешних факторов. Правда, компоненты вектора θ , за некоторым исключением, не допускают несмещённых оценок, но огромные объёмы планов (несколько тысяч циклов в час), свидетельствуют о возможности успешного использования асимптотических методов получения оценок.

АВТОМАТИЗИРОВАННАЯ МЕТОДИКА ВЫДЕЛЕНИЯ ЕСТЕСТВЕННЫХ ГИДРОЛОГИЧЕСКИХ АНОМАЛИЙ В ЦЕЛЯХ ЭКОЛОГИЧЕСКОГО МОНИТОРИНГА

С.Т.Трусенков, О.О.Трусенкова

С усилением влияния человека на природную среду становится все необходимее критическое сравнение количественных характеристик этого влияния с интенсивностью естественных вариаций. Это, в частности, касается гидрологических параметров вод морей и океанов, являющихся индикатором теплового и химического загрязнения. Объем эмпирических сведений о естественных аномалиях (ЕА) и, следовательно, оценки критичности экологической ситуации зависят от методики выделения ЕА.

Естественными гидрологическими аномалиями следует считать отклонения от климатического "фона", определяемого как среднее по статистическому ансамблю климатических состояний. В качестве "фона" не следует брать лишь традиционные климатические средние, а необходимо определять его путем выделения естественных климатических районов и естественных гидрологических периодов.

Выделению и описанию вероятностных свойств ЕА служит единая автоматизированная методика поиска многомерных статистических зависимостей между океанографическими полями на больших акваториях, различных горизонтах, на больших временных интервалах [1, 2]. Она обеспечивает:

- 1) изучение многомерных связей между задаваемыми признаками и вычисление информативных признаков для эффективного описания сходства океанографических объектов;
- 2) классификацию объектов в многомерном пространстве информативных признаков по устанавливаемому критерию сходства;
- 3) восстановление эмпирической плотности распределения отображением результатов вышеуказанных операций в географическое пространство (в виде районирования) и на временную ось (в виде периодизации полей).

Многомерный статистический анализ наблюдений организован

в поисковом режиме гибкого манипулирования обрабатываемыми выборками и техниками использования базовых вычислительных алгоритмов: корреляционного, факторного, спектрального, кластерного анализа. Это достигается за счет привлечения языковых средств описания задачи и управления промежуточной статистической информацией, графических средств, облегчающих неформальную интерпретацию и диалоговых средств, позволяющих принимать эвристические решения.

Методика реализована в виде пакета прикладных программ ДИСТРИКТ, оснащенного собственным архивом данных и имеющего программный интерфейс с базой данных Регионального океанографического банка по Тихому и Индийскому океанам.

На этой основе построены информативные зоны температуры поверхности (ТПО) в северной части Тихого океана, что позволяет более обоснованным образом выделять климатический "фон". В частности, подтверждена особая роль энергоактивной зоны Куроисио в формировании регулярных колебаний ТПО и малая активность в этом отношении центральноокеанической области. Типизированы сезонные формы пространственных связей поверхностных аномалий, выделенных различными способами. Проанализирована сезонная изменчивость температуры деятельного слоя Куроисио, в результате чего выявлены наиболее информативные водные слои и естественные гидрологические периоды. Это позволило оценить несезонные температурные аномалии. Получен ряд выводов, существенных для изучения чувствительности среды к внешним возмущениям, что, в частности, немаловажно для экологических прогнозов.

Литература

1. Трусенков С.Т., Трусенкова О.О. Анализ геофизических полей на основе многомерных статистических методов: Препринт. Владивосток: ИАПУ ДВНЦ АН СССР. 1984. 45 с.
2. Трусенкова О.О., Трусенков С.Т. Пакет прикладных программ для климатического анализа океанографических полей: Препринт. Владивосток: ИАПУ ДВО АН СССР. 1988. 36 с.

СТАТИСТИЧЕСКАЯ КЛАССИФИКАЦИЯ УРОВНЕЙ ВНЕЗАПНОЙ СМЕРТНОСТИ В ЗАВИСИМОСТИ ОТ МЕТЕОФАКТОРОВ

Г.Л.Шевляков

В течение трех лет в г.Архангельске собиралась статистика ежедневной внезапной смертности по причинам сердечно-сосудистой патологии. Вместе с числом внезапных смертей ежедневно фиксировались значения следующих метеофакторов: градиента атмосферного давления, скорости ветра, среднесуточной температуры и ее градиента, уровня активности земного магнитного поля (АК-индекс) и ряда факторов, описывающих солнечную активность (площадь солнечных пятен, число Вольфа, AS -индекс и др.).

Ставилась задача выявления наиболее опасных в смысле риска внезапной смерти метеофакторов, их комбинаций, а также порогов "риска" для них.

Для решения этой задачи использовались такие статистические процедуры многомерного анализа, как дискриминантный и факторный анализ, причем в классическом и робастном вариантах. Робастные модификации классических процедур включали в себя робастные оценки математических ожиданий, дисперсий и элементов ковариационных и корреляционных матриц [1,2]. Необходимость применения робастных методов обуславливалась наличием в исходных данных выбросов, связанных, например, с резким увеличением уровня внезапной смертности в определенные дни при вполне благоприятных метеоусловиях.

В докладе приводятся результаты решения описанной задачи различными методами и обсуждаются полученные зависимости уровня внезапной смертности от значений метеофакторов.

Л и т е р а т у р а

1. Хьюбер П. Робастность в статистике. - М.: Мир, 1984.
2. Пасман В.Р., Шевляков Г.Л. Робастные методы оценивания коэффициента корреляции. - Автоматика и телемеханика, 1987, № 3, с.70-80.

КОМПЛЕКСНЫЙ МНОГОМЕРНЫЙ АНАЛИЗ ЭФФЕКТИВНОСТИ ОБСЛУЖИВАНИЯ ОБОРУДОВАНИЯ

Н.В.Алтухова

Влияние факторов, составляющих систему статистических показателей, оказывающих воздействие на уровень обслуживания оборудования с целью повышения его эффективности исследовалось с помощью комплексного применения факторного анализа, регрессии на факторах и последующего кластерного анализа.

Исходную совокупность составило 101 предприятие машиностроительной промышленности. При обработке материалов использованы пакеты прикладных программ, разработанные РВЦ ЦСУ РСФСР.

В качестве первоначальных факторов, влияющих на эффективность организации обслуживания оборудования были отобраны показатели, отражающие: а/ технический уровень – среднегодовая стоимость производственного оборудования, средний возраст производственного оборудования, средний возраст металлорежущего оборудования, коэффициент сменности производственного оборудования, коэффициент сменности в ремонтных цехах, техническая нагрузка на одного рабочего, занятого обслуживанием оборудования; б/ экономический уровень – средняя выработка продукции на одного основного рабочего; средняя выработка на одного рабочего, занятого обслуживанием оборудования; затраты на капитальный ремонт оборудования, доля простоев из-за неисправностей оборудования и его обслуживания, затраты заработной платы промышленно-производственного персонала на рубль продукции; в/ организационный уровень – коэффициент сменности рабочих, занятых обслуживанием оборудования; средний процент выполнения норм выработки рабочими, количество технически обоснованных норм, количество специализированных или комплексных бригад, занятых обслуживанием оборудования; число рабочих, охваченных мероприятиями НОТ, направленными на улучшение обслуживания оборудования; фактические затраты по внедрению этих мероприятий НОТ. Выбор показателей определял-

ся результатами многофакторного регрессионного анализа.

Методом факторного анализа была построена модель, подтверждавшая правомерность разделения процесса обслуживания оборудования на три аспекта: технический, экономический и организационный. При этом для вычисления матрицы коэффициентов парной корреляции применялся коэффициент Пирсона, оценки факторных нагрузок определялись методом максимального правдоподобия, в качестве начальных приближений использовались нагрузки, полученные центроидным методом. В расчете проверялись гипотезы о значимом числе обобщенных факторов 2,3 и 4 соответственно с 134,117 и 101 степенями свободы. На базе полученных в результате факторного анализа значений обобщенных факторов получено уравнение множественной регрессии:

$$y = 19,2 + 515,97f_1 - 340,61f_2 - 607,37f_3 .$$

Все коэффициенты регрессии оказались статистически значимыми при проверке по t -критерию для 5%-ного уровня значимости. Знаки при коэффициентах регрессии не противоречат экономическому содержанию обобщенных факторов. Уравнение регрессии проверялось с помощью критерия Фишера, значение которого свидетельствует о существенности уравнения для 5%-ного уровня значимости. Обобщенные факторы обуславливают 77,1% вариации коэффициента эффективности организации обслуживания оборудования.

С помощью кластерного анализа все предприятия указанной совокупности были сгруппированы в зависимости от полученных обобщенных факторов. Выбор кластерного анализа обусловлен тем, что он имеет все преимущества комбинационной группировки, но свободен от главного недостатка - распыления материала. Число кластеров определялось в процессе разбиения множества. Число группировочных признаков было рассчитано и принято равным четырем.

Классификация предприятий с использованием кластерного анализа позволила ранжировать предприятия на отстающие, средние и передовые в смысле организации обслуживания оборудования, как по отдельным подгруппам, так и по всей совокупности в зависимости от любого показателя, включенного в схему анализа. Определены направления целесообразного воздействия на каждый из показателей, характеризующих технический, экономический и организационный уровень обслуживания оборудования.

НЕЧЕТКИЕ МНОЖЕСТВА И НЕЧЕТКАЯ СТАТИСТИКА В ЗАДАЧАХ ОЦЕНКИ КАЧЕСТВА

А.П.Демиденко

Задачу оценки качества продукции можно рассматривать в виде задачи анализа данных, связанную с обработкой информации и ее сжатием. При этом качество продукции характеризуется вектором показателей, элементы которого включают числовые и нечисловые характеристики. Предлагается решать эту задачу путем свертки вектора показателей $x = (x_1, \dots, x_i, \dots, x_n)$ в скаляр y , применяя для ее решения идеи классификации. При этом оцениваемое состояние \mathcal{B} , характеризуемое вектором x , необходимо отнести к одному из классов ℓ , $\ell = \overline{1, M}$. Здесь M - число градаций шкалы $\mathcal{Y}_y = \{\alpha_1, \alpha_n\}$, в которой измеряется показатель y . С этой целью зависимость определенной степени проявления свойства, описываемого показателем y , например α_ℓ , от значений показателя x_i выразим в виде нечеткого множества $\mathcal{A}_\ell(x_i) = \{\beta, \mu_{\mathcal{A}_\ell}(\beta)\}$. Здесь $\beta \in \mathcal{Y}_{x_i}$. \mathcal{Y}_{x_i} - шкала измерения показателя x_i . Функцию принадлежности $\mu_{\mathcal{A}_\ell}(\beta)$ можно рассматривать в качестве меры степени соответствия β -го значения показателя x_i значению α_ℓ показателя y . При решении задачи классификации предполагается, что для каждого класса ℓ (оценки $\alpha_\ell \in \mathcal{Y}_y$) на декартовом произведении $Q = \mathcal{Y}_{x_1} \times \dots \times \mathcal{Y}_{x_i} \times \dots \times \mathcal{Y}_{x_n}$, представляющем собой множество возможных состояний показателей x_i , существует идеальная точка A_ℓ , т.е. наиболее типичное состояние, которое можно принять за центр класса. Оцениваемое состояние относят к тому классу, близость до центра которого максимальна. Рассмотрены алгоритмы оценки для случаев, когда показатели x_i заданы точными значениями, нечеткими множествами, измеряются в двухальтернативной шкале. Использование нечеткой меры, характеризующий субъективное представление о возможности оценки α_ℓ , $\forall \ell = \overline{1, N}$, позволяет определить наиболее типичную оценку состояния \mathcal{B} с помощью нечеткого интеграла. Рассматриваются вопросы идентификации нечеткой меры. Предлагаемый подход применен в системе управления качеством.

ПРИМЕНЕНИЕ МЕТОДОВ АНАЛИЗА СТАБИЛЬНОСТИ ЗНАЧЕНИЙ ИНФОРМАТИВНЫХ ПАРАМЕТРОВ ДЛЯ ОЦЕНКИ КАЧЕСТВА ЭЛЕКТРОРАДИОИЗДЕЛИЙ

В.Г.Кобзев, Н.Г.Николенко, А.А.Репко

Перспективным направлением работ в области автоматизации контроля и испытаний электрорадиоизделий (ЭРИ) является разработка статистических методов обработки результатов неразрушающего контроля, который обеспечивает проведение необходимых измерений параметров или показателей качества изделий без ухудшения присущих им свойств. Неразрушающий характер производственных испытаний предполагает контроль состояния ЭРИ по их специальным информативным параметрам, анализ совокупностей которых демонстрирует их существенную статистическую неоднородность и нестабильность. Задача контроля в этом случае ставится как статистическая задача принятия решения об одном из альтернативных состояний ЭРИ (т.е. оценивания их качества) на основе анализа статистических свойств значений их параметров.

Анализу подвергаются как статическая ситуация, так и динамика изменения информативных параметров. В первом случае совокупности значений информативных параметров группы изделий анализируются на наличие подгрупп или отдельных аномальных значений путем использования процедур разделения смеси и тестирования выбросов. Во втором случае исследуется специальный показатель "подвижности" информативного параметра, характеризующий скорость изменения его во времени у конкретного изделия. Редкие значения параметров, а также их резкие изменения во времени характеризуют нестабильность свойств испытываемых изделий, что свидетельствует о недостаточно высоком уровне их качества.

Указанный подход представляет собой один из способов формализации задачи оценивания качества изделий при наличии соответствующей шкалы. Его применение позволяет решить задачу выявления потенциально-ненадежных ЭРИ по результатам их испытаний.

МЕТОД ВЫДЕЛЕНИЯ ЕДИНИЦ БЕСФОРМЕННОЙ ПРОДУКЦИИ ПРИ СТАТИСТИЧЕСКОЙ КОНТРОЛЕ КАЧЕСТВА

Г.Г.Кравченко, А.И.Орлов

Методы статистического контроля ориентированы на штучную продукцию, единицы которой физически отделены друг от друга. В различных отраслях народного хозяйства применяется бесформенная продукция - порошкообразная, жидкая, газообразная и др. В традиционных методах статистического контроля предполагается, что результаты измерения контролируемого параметра для различных единиц продукции можно моделировать как независимые случайные величины. Для бесформенной продукции результаты измерений для проб, взятых в соседних точках, коррелируют между собой, а для взятых в далеких точках, - как правило, нет. Предлагаемый метод основан на выделении минимального расстояния d_0 такого, что при взятии проб в точках, расстояние между которыми больше d_0 , коэффициент корреляции между результатами измерений незначимо отличается от 0. Затем трехмерный объем, занятый бесформенной продукцией, предлагается представить в виде объединения тел (например, кубов), внутрь каждого из которых вписан шар радиуса d_0 . Каждое такое тело предлагается принять за единицу продукции, а измерение проводить в центре соответствующего шара. Тогда задача контроля качества бесформенной продукции сводится к задаче для штучной продукции, для которой методы статистического контроля хорошо разработаны.

Для экспериментального определения d_0 предлагается при различных d анализировать результаты измерений в точках A_i, B_i , $i = 1, 2, \dots, n$, таких, что расстояние между A_i и B_i равно d , а расстояние между различными парами точек (A_i, B_i) много больше d . При контроле плотности баритового утяжелителя, предназначенного для регулирования плотности буровых растворов, предлагаемый метод с использованием коэффициента ранговой корреляции Спирмена ($n = 10$, $\alpha = 0,1$) дает $d_0 = 25$ см.

ПРИМЕНЕНИЕ МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА ДЛЯ РАНЖИРОВАНИЯ ПРИЧИН НАРУШЕНИЙ КАЧЕСТВА РАБОТЫ КОМБАЙНОВ

Т.К.Курочкина

При настройке зерноуборочных комбайнов обычно сложно определять причины нарушений качества, что приводит к увеличению потерь зерна. Применение многомерного статистического анализа для оценки нарушений качества позволяет ранжировать причины нарушений. Полученные ранги используются при разработке способа настройки комбайнов.

Работа комбайна оценивается качественными показателями K_i потерями, дроблением и засоренностью зерна. Если хотя бы один из показателей больше нормированного значения, то наблюдается нарушение качества работы комбайнов.

Нарушения качества вызваны причинами - отклонениями регулируемых параметров комбайна от оптимального значения.

Оценка показателей K_i выполнена по методу Мизес-Алимова, при котором рассчитывается заданный параметр распределения изучаемой величины для нарастающего объема выборки в целях устойчивости оценки параметров. Оценка показателя определяется интервалом устойчивости, заданным последними значениями оценки в предыдущих сериях испытаний.

Собрана статистика нарушений качества работы комбайнов и причин их вызывающих. Ранжирование причин нарушений качества выполнено по методу *Kitter* квазиоптимального ранжирования факторов процесса. /1,2/.

Значения рангов причин нарушений качества рассчитаны по программе, написанной на языке Фортран 4 и использованы при разработке способа и средства настройки комбайнов, которые прошли производственную проверку в хозяйствах Новосибирской области.

ЛИТЕРАТУРА

1. Кутин Г.И. Методы ранжировки комплексов признаков //Зарубежная радиоэлектроника.- 1981.- №9.- стр.54-70.
2. *Kitter J. On nonlinear metric selection for feature selection in measurement space. // Inform. Sciences.- 1978.- V.9.- p. 359-363.*

НЕПАРАМЕТРИЧЕСКИЙ АЛГОРИТМ ОЦЕНКИ КАЧЕСТВА ПРОДУКЦИИ ПРИ ИЗМЕНЯЮЩИХСЯ УСЛОВИЯХ

А.В.Медведев, И.А.Пестунов

Оценка (прогноз) качества продукции по косвенным показателям является типичной задачей для многих промышленных предприятий. С математической точки зрения эта задача часто рассматривается в рамках известных постановок задач распознавания образов [1]. Однако на практике "истинная" разделяющая поверхность в пространстве контролируемых переменных, характеризующих то или иное состояние объекта, как правило дрейфует во времени заранее непредвиденным образом. В настоящем докладе излагается непараметрический алгоритм классификации, основанный на парзеновских оценках многомерной плотности распределения, и результаты его исследования для подобных условий.

Рассмотрим следующую постановку задачи. Пусть в евклидовом пространстве R^k с вероятностями q_1, q_2 ($q_1 + q_2 = 1$) регистрируются наблюдения из классов ω_1, ω_2 соответственно. Наблюдение из $\omega_i, i=1,2$ представляет собой реализацию случайного вектора $\xi_i \in R^k$, плотность распределения которого $P_s(x/\omega_i)$, где $x \in R^k$, является неизвестной функцией, зависящей от времени S . Имеется обучающая выборка (ОВ) $X = X^1 \cup X^2$ объема $N = N_1 + N_2$, где $X^i = \{x_t^i: x_t^i - \text{случайное наблюдение из } \omega_i \text{ в дискретный момент времени } t, t = \overline{1, N_i}\}$. Требуется построить алгоритм для классификации случайных наблюдений из R^k , учитывающий временной дрейф вероятностных характеристик классов (ВХК).

Ясно, что с практической точки зрения такая постановка может быть оправдана только при условии, если скорость изменения ВХК значительно меньше темпа поступления обучающей информации.

Для конструирования непараметрического алгоритма распознавания, предназначенного для работы в нестационарных условиях, используется известный прием [2], суть которого состо-

ит в том, что при оценивании плотности $p_s(x/\omega_i)$ вновь поступающая информация получает больший вес, по сравнению со "старой". Предлагаемый алгоритм записывается следующим образом:

$$\left. \begin{array}{l} x \in \omega_1 \quad , \text{ если } q_1 \hat{p}_s(x/\omega_1) > q_2 \hat{p}_s(x/\omega_2) \\ x \in \omega_2 \quad - \text{ в противном случае,} \end{array} \right\} \quad (I)$$

где
$$\hat{p}_s(x/\omega_i) = N_i^{-1} \sum_{t=1}^{M_i} c^{-K} \Phi((x - x_t^i)/c) \rho(M_i - t);$$

$c^{-K} \Phi((x - x_t^i)/c)$ — колоколообразная функция, удовлетворяющая определенным условиям [1]; c — параметр сглаживания; $\rho(M_i - t)$ — функция памяти, убывающая с увеличением аргумента.

В силу временного дрейфа разделяющей поверхности нельзя говорить о какой-либо асимптотической сходимости алгоритма, но в докладе приводятся результаты его численного исследования для различных значений параметра, определяющего скорость изменения ВХК, и параметра Δt , характеризующего темп поступления обучающей информации. Как и следовало ожидать, при фиксированном Δt с увеличением скорости изменения ВХК качество распознавания ухудшается и наоборот. Если изменчивость характеристик незначительна, то классификатор (I) работает удовлетворительно, в то время как обычные алгоритмы оказываются неработоспособными. Рассматриваются также алгоритмы корректировки ОБ с учетом временного дрейфа ВХК и вопросы оптимизации классификатора (I).

Л и т е р а т у р а

1. Живоглазов В.И., Медведев А.В. Непараметрические алгоритмы адаптации. Фрунзе, Илим, 1974.
2. Фельдбаум А.А. Основы теории оптимальных автоматических систем. М., Наука, 1962.

МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ КАЧЕСТВА И НАДЕЖНОСТИ ПРОКАТНЫХ ВАЛКОВ

В.С.Мхитарян, С.В.Чуприн

Удовлетворение потребностей народного хозяйства в листовом металле во многом зависит от уровня качества и надежности важнейшего сменного инструмента металлургического оборудования – прокатных валков (ПВ). Затраты на изготовление и ремонт ПВ за время эксплуатации прокатного стана в несколько раз превышают его стоимость.

Исследованы законы распределения отказов и времени безотказной работы ПВ. Построены модели постепенных, внезапных и усталостных отказов с учетом особенности эксплуатации валков в различных клетях прокатного стана. Анализ показал, что на увеличение ресурса и надежности функционирования ПВ большое влияние оказывает их правильная и рациональная эксплуатация, маршруты движения по клетям стана, оптимальный режим перевалок и перешлифовок при восстановлении работоспособности.

При построении статистической модели рабочего слоя прокатных валков выявлены недостатки в технологии их эксплуатации, связанные с неравномерностью расхода рабочего слоя по клетям.

Разработаны рекомендации по упорядочению числа установок валков по различным клетям прокатного стана в зависимости от величины рабочего слоя и скорости его расхода.

Построены множественные регрессионные модели ресурса и наработки до отказа ПВ как по исходным показателям производственного и эксплуатационного качества, так и по главным компонентам, проанализировано влияние главных компонент, характеризующих качество восстановления работоспособности, состояние рабочего слоя и эффективность использования ПВ.

ПРОЕКТ "СТАТПРОМ"

А.И.Орлов, О.М.Черномордик

В 1985 г. была создана Рабочая группа по упорядочению системы стандартов по прикладной статистике и другим статистическим методам из 56 специалистов академических институтов, вузов (МГУ, МЭИ, МВТУ и др.), НИИ Госстандарта СССР, организаций и предприятий отраслей народного хозяйства, в том числе здравоохранения (НИИ гигиены труда и профзаболеваний, Институт хирургии им.А.В.Вишневского и др.). Она проанализировала 31 государственный стандарт по статистическим методам. По рекомендациям Рабочей группы уже отменено 24 стандарта. В 1988 г. принято решение о разработке аванпроекта комплекса методических документов и пакетов программ по статистическим методам стандартизации и управления качеством. Для выполнения этой работы советско-франко-итальянское совместное предприятие ИНТЕРКВАДРО по заданию ВНИИ стандартизации Госстандарта СССР организовало коллектив из 25 специалистов 11 организаций (в том числе 9 докторов и 13 кандидатов наук). В основу разработки проекта, названного СТАТПРОМ (статистика в промышленности), положены следующие принципы.

1. Ориентация на непрофессионального пользователя статистических методов – инженера промышленного предприятия, сотрудника отраслевого НИИ: технического, медицинского и др.

2. Системный подход с целью охвата всего многообразия имеющихся статистических методов. Он предполагает проведение большого объема методологических работ с целью анализа и упорядочения статистических методов, обоснованного выбора алгоритмов для решения тех или иных задач. На допрограммной стадии разработки проекта необходимы и теоретические исследования в области математической статистики.

3. Обеспечение современного научно-технического уровня. Включен ряд последних разработок советской вероятностно-

ли доктора наук по соответствующим областям. Это позволяет надеяться на конкурентноспособность результатов разработок на мировом рынке пакетов статистических программ.

4. Пакеты программ должны сопровождаться методическими документами, содержащими описание методов (включая расчетные формулы алгоритмов), их краткое теоретическое обоснование (со ссылками на публикации), подробное описание примеров применения в конкретных прикладных областях (формулировки задач в содержательных терминах, примеры обработки реальных данных).

5. Ориентация на имеющуюся в СССР технику – как ИБМ-совместимую, так и ДЕК-совместимую (ЭВМ серий СМ, ДБК, Электроника, Мера и др.), прежде всего на ПЭВМ.

6. Большое внимание вопросам распространения, внедрения, сопровождения методических документов и пакетов программ, обучения специалистов основам статистических методов, необходимым для их осознанного применения, прежде всего "входам" и "выходам" алгоритмов.

Аванпроект содержит следующие разделы: развитие, стандартизация и внедрение статистических методов управления качеством продукции: анализ состояния в СССР и за рубежом; термины и определения в области статистических методов; методологические и методические вопросы современной прикладной статистики и других статистических методов; прикладная статистика (с разделением на статистику случайных величин, многомерный статистический анализ, статистику временных рядов и случайных процессов, статистику объектов нечисловой природы), статистический контроль качества продукции, статистическое регулирование технологических процессов (в том числе с помощью контрольных карт), вопросы внедрения статистических методов, программного и экономического обеспечения. Общ. объем аванпроекта – более 1500 стр.

В качестве прототипов программных продуктов по проекту СТАТИПРОМ можно рассматривать: по прикладной статистике – пакет "Анализ данных" (НИИ гигиены труда и профзаболеваний), по статистическому контролю – разработки Горьковского филиала ВНИИНМАШ Госстандарта СССР, по статистическому регулированию технологических процессов – МЭИ (Г.Ф. Филаретов).

АНАЛИЗ ДАННЫХ И ИНТЕГРАЛЬНЫЕ ОЦЕНКИ КАЧЕСТВА ПРИ ПРОЕКТИРОВАНИИ СЛОЖНЫХ СИСТЕМ

И.Б.Руссман, В.М.Умывакин

Выделение парето-оптимальных вариантов является "ядром" современных методик оценки качества продукции. Преимущество таких методик заключается в возможности выбора наилучших вариантов на основе объективной информации. Однако, их работоспособность существенно снижается при наличии большого числа противоречивых показателей. Кроме того, парето-оптимальные варианты обладают значительным разнообразием оцениваемых свойств. Выбор среди них наилучших вариантов возможен лишь на основе всестороннего анализа множества Парето.

Предлагается процедура анализа данных, позволяющая формировать интегральные оценки качества парето-оптимальных вариантов при проектировании сложных систем. Она основана на традиционной схеме классификации множества Парето (последовательности типологических группировок, задаваемых наиболее существенными показателями) и оригинальных способах построения непрерывных функций ценности внутри каждого класса (типа), которые являются интегральными оценками качества. Отметим, что при традиционном способе классификации некоторые варианты могут принадлежать сразу нескольким типам, а различные по совокупности свойств варианты могут быть отнесены к одному типу и в результате получить одинаковые оценки качества.

Варианты, отнесенные к каждому типу, образуют обучающую выборку для построения функций ценности, которые принимают приблизительно одинаковые значения для вариантов из одного класса и различные для вариантов из разных типов. Они формируются на основе информации об эквивалентности и предпочтительности оцениваемых вариантов. Эта информация задается на классах вариантов, т.е. предполагается известным количество типов и их ранжирование по качеству.

ОБ ОЦЕНКЕ ВЛИЯНИЯ СТАБИЛИЗАЦИИ ФАКТОРОВ НА КАЧЕСТВО ПРОДУКЦИИ

А.П.Рыжаков, А.Ф.Терпугов

Пусть качество продукции определяется величиной y , нулевое значение которой означает идеальное соответствие стандарту. Будем считать, что y представима в виде

$$y = \sum_{i=1}^k a_i x_i + n = x^T a + n \quad (I)$$

где величины x_i доступны измерению, а величина n представляет собой неконтролируемое воздействие. Будем считать, что совокупность величин x_1, \dots, x_k является k -мерной нормальной случайной величиной с вектором средних $\mu = 0$ и ковариационной матрицей Σ . Величину n будем считать нормальной случайной величиной с $Mn = 0$ и $D\{n\} = \sigma_n^2$. Тогда $My = 0$, а дисперсия имеет вид

$$D\{y\} = \sigma_n^2 + a^T \Sigma a \quad (2)$$

Пусть теперь какая-то часть факторов x (например, x_1, \dots, x_e) застabilизирована. Для удобства будем использовать следующие обозначения. Вектор R размерности k будем представлять в блочном виде:

$$R = \begin{bmatrix} R^{(1)} \\ R^{(2)} \end{bmatrix}; R^{(1)} = \begin{bmatrix} x_1 \\ \vdots \\ x_e \end{bmatrix}, R^{(2)} = \begin{bmatrix} x_{e+1} \\ \vdots \\ x_k \end{bmatrix}.$$

Матрицу C размерности $k \times k$ будем записывать в блочном виде:

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}; C_{11} = \begin{bmatrix} c_{11} & \dots & c_{1e} \\ \vdots & \ddots & \vdots \\ c_{e1} & \dots & c_{ee} \end{bmatrix}, C_{12} = \begin{bmatrix} c_{1,e+1} & \dots & c_{1k} \\ \vdots & \ddots & \vdots \\ c_{e,e+1} & \dots & c_{ek} \end{bmatrix},$$

$$C_{21} = \begin{bmatrix} c_{e+1,1} & \dots & c_{e+1,e} \\ \vdots & \ddots & \vdots \\ c_{k1} & \dots & c_{ke} \end{bmatrix}, C_{22} = \begin{bmatrix} c_{e+1,e+1} & \dots & c_{e+1,k} \\ \vdots & \ddots & \vdots \\ c_{k,e+1} & \dots & c_{kk} \end{bmatrix}$$

Оставшиеся факторы x_{e+1}, \dots, x_k образуют $(k-e)$ -мерную нормальную случайную величину с вектором средних $\mu^{(2)} = 0$ и ковариационной матрицей $[1]$:

$$z = \sum_{12} - \sum_{21} \sum_{11}^{-1} \sum_{12} \quad (3)$$

Тогда дисперсия y станет равной

$$D\{y/x^{(1)}\} = D\{n\} + a^{(2)T} z a^{(2)} \quad (4)$$

Очевидно, что от стабилизации x_1, \dots, x_e дисперсия умень-

шается и второе слагаемое (4) есть та остаточная дисперсия, которая обусловлена оставшимися переменными x_{k1}, \dots, x_{kN} .

Величины $\sigma_k^2, \alpha, \Sigma$ неизвестны, поэтому воспользоваться (2) - (4) можно лишь оценив слагаемые вида $\sigma_k^2, \alpha^T \Sigma \alpha, \alpha^{(i)T} \alpha^{(i)}$ по результатам наблюдений.

Пусть проведено N опытов и в j -ом опыте наблюдались величины y_j и $x_i^{(j)}, i = \overline{1, K}$, связанные соотношением

$$y_j = \sum_{i=1}^K a_i x_i^{(j)} + n_j \quad (j = \overline{1, N})$$

Обозначим

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}, X = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_K^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \dots & x_K^{(N)} \end{bmatrix}$$

Можно показать, что в качестве несмещенной оценки величины

$$\sigma_k^2 \text{ можно взять величину } S_k^2 = \frac{1}{N-K} (Y - X(X^T X)^{-1} X^T Y)^T (Y - X(X^T X)^{-1} X^T Y) \quad (5)$$

Несмещенной оценкой величины $\alpha^T \Sigma \alpha$ является

$$S_\alpha^2 = \frac{1}{N} [Y^T X (X^T X)^{-1} X^T Y - K \cdot S_k^2] \quad (6)$$

В качестве оценки величины $\alpha^{(i)T} \alpha^{(i)}$ в (4) следует взять

$$S_{\alpha^{(i)}}^2 = \frac{1}{N} \{ (\hat{\alpha}^{(i)})^T [C_{\alpha\alpha} - C_{\alpha 1} \cdot C_{11}^{-1} \cdot C_{1\alpha}] \hat{\alpha}^{(i)} - (K-1) S_k^2 \} \quad (7)$$

где $\hat{\alpha} = C^{-1} X^T Y, C = X^T X$.

Несмещенная оценка $S_{\alpha^{(i)}}^2$ остаточной дисперсии y при стабилизации ℓ факторов может быть вычислена для любого подмножества факторов X и характеризует степень влияния разброса значений этих факторов на качество продукции. Перебор подмножеств факторов и сравнение остаточных дисперсий позволит выявить набор факторов, стабилизация значений которых существенно улучшает качество продукции.

ЛИТЕРАТУРА

- I. Siotani, Hayakawa. *Modern multivariate statistical Analysis*. Columbus, Ohio: Amer. Sci. Press, 1985.

НОВОЕ В ОЦЕНКЕ КАЧЕСТВА ТЕКСТИЛЬНЫХ МАТЕРИАЛОВ

С.В.Соболев

Работа ориентирована на компьютеризацию текстильных лабораторий, а в основу ее положена концепция неоднородных статистических совокупностей. Для различных текстильных материалов выявлены механизмы возникновения неоднородных совокупностей, описаны их типы. Разработаны методы, алгоритмы и программы анализа таких совокупностей. Выбор того или иного метода зависит от характера и объема выборки, сути решаемой задачи. Большая часть эмпирических распределений аппроксимируется 3-4 -х параметрическими кривыми: модель Тькки, составное симметричное распределение, кривые Джонсона. При испытании прочности материалов со слабыми участками разных типов (смешанная пряжа, ткани из нее, химические и углеродные волокна и др.) 4-х параметрических кривых недостаточно.

Важную технологическую характеристику хлопка - долю слабых волокон - простым подсчетом надежно оценить не удавалось. Аппроксимацией распределения показано, что волокна хлопка даже одного сорта, но снятые с разных конусов цветения, по содержанию волокон с разрывной нагрузкой $< I$ и < 2 сН отличаются в 1.5 раза.

При определении длины хлопковых волокон механизированным способом они рвутся, причем в разных случаях в различной степени. Оценка доли коротких волокон в испытуемой партии настолько смещена и неэффективна, что этот важнейший показатель качества хлопка-волокна до сих пор не стандартизован. Смещение и неэффективность устранены аппроксимацией плотности распределения по усеченным сгруппированным данным.

Показано, что обращение к одному из классических распределений катастрофически искажает оценку вероятности появления слабых мест в пряже. Исследуя массивы данных, со-

держащие тысячи чисел, можно сопоставлять по прочности пряжу разных фабрик.

Построена математическая модель многократных разрывающих усилий, действующих на нить основы в ткачестве. Согласно модели уровень обрывности существенно зависит от разброса амплитуды пиков натяжения, то есть от состояния оборудования.

Статистическая модель возникновения обрыва по причине "слабая нить" в ткачестве учитывает физико-механические характеристики пряжи, а также конструктивные особенности ткацких станков, определяющие характер воздействия на нить. Модель дает возможность сделать новый вывод о том, что значительного запаса по прочности пряжи не имеет.

Частоту возникновения обрывов, а также дефекты ткани лучше описывает смесь распределений Пуассона, чем отрицательное биномиальное распределение; нагляднее становится интерпретация данных.

Эталонный способ измерения длины куска ткани – вручную. Разрешается использовать механизированные средства, если ошибка $\leq 0,3\%$. Построены регрессионные зависимости для вычисления поправок в зависимости от плотности и веса ткани, от состава нитей. После внесения поправки ошибка не превышает заданной границы, это открывает более широкие возможности для механизации ручной операции.

Для достижения высокой прочности композитов армирующие волокна должны иметь близкие кривые деформации. Введено понятие статистической неоднородности материала. Полезной для прогноза прочности углепластиков оказалась такая мера статистической неоднородности, как χ^2 – расстояние между эмпирическим распределением и аппроксимирующей 4-х параметрической кривой.

Изучена достоверность расчетов среднего и дисперсии вручную путем предварительной группировки. Частота появления ошибок лаборанта намного выше, чем при более простых операциях с данными.

ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ ПОКАЗАТЕЛЕЙ НАДЕЖНОСТИ ПО РЕЗУЛЬТАТАМ ФОРСИРОВАННЫХ ИСПЫТАНИЙ

О.И.Тескин, С.Я.Криволапов

Одной из важнейших задач при контроле качества и надежности промышленной продукции является сокращение продолжительности и объема испытаний за счет использования форсированных режимов. Этой теме посвящено большое число работ [1], однако многие вопросы пока остаются открытыми. В частности, не разработаны алгоритмы интервального оценивания показателей надежности изделий по результатам форсированных испытаний. В докладе представлено аналитическое решение задачи интервального оценивания показателя безотказности $R(x, t)$, представляющего собой вероятность безотказной работы (БЕР) технического изделия за время t при действии нагрузки $x \in E_0 \subset R^1$, где E_0 — множество допустимых значений x , при следующих предположениях.

1. Функции распределения наработки до отказа изделия при различных значениях нагрузки x совпадают с точностью до преобразования масштаба (линейная модель).

2. Семейство функций распределения наработки до отказа изделий принадлежит классу ВСФИ [2], т.е. классу распределений с возрастающей в среднем функцией интенсивности отказов.

Обозначим через $T(x)$ наработку до отказа в режиме x .

Первое предположение согласно [3] влечет существование положительной возрастающей функции $\gamma(x)$, такой, что $\forall x \in E_0$ распределение случайной величины $\gamma(x)T(x)$ не зависит от значения x . Это требование эквивалентно тому, что с изменением режима испытаний изменяется лишь масштабный параметр семейства распределений $F(x, t) = 1 - R(x, t)$, т.е.

$$R(y, t) = R\left(x, \frac{\gamma(y)}{\gamma(x)} \cdot t\right), \quad \forall x, y \in E_0 \quad (1)$$

Функция $\gamma(x)$ имеет смысл скорости исчерпания ресурса при действии нагрузки x и ее вид определяется моделью форсирования.

Второе допущение означает, что $\forall t_1, t_2, t_1 < t_2$ справедливо неравенство [2]

$$R(x, t_1) \geq [R(x, t_2)]^{t_1/t_2}, \quad x \in E_0. \quad (2)$$

Основной результат состоит в следующем. Пусть $\underline{R}_\mu(x, t)$ есть нижняя доверительная граница (НДГ) уровня μ для ВБР $R(x, t)$, полученная по результатам биномиальных испытаний для заданных значений x и t . Тогда справедлива

Теорема. Если по результатам испытаний на промежутке $[0, t_2]$ при нагрузке $x_2 > x_1$ найдена НДГ уровня μ $\underline{R}_\mu(x_2, t_2)$ для $R(x_2, t_2)$, то выражение НДГ $\underline{R}_\mu(x, t_1)$ для ВБР $R(x, t_1)$ на промежутке $[0, t_1]$ при нагрузке $x_1 (t_1 > t_2)$ имеет вид:

$$\underline{R}_\mu(x_1, t_1) = [\underline{R}_\mu(x_2, t_2)]^{\gamma}, \quad \gamma = \gamma(x_1)t_1 / \gamma(x_2)t_2. \quad (3)$$

Формула (3) дает правило интервального оценивания ВБР изделия в нормальном режиме x_1 за время t_1 , если найдена НДГ для ВБР в форсированном режиме $x_2 > x_1$.

Л и т е р а т у р а

1. Каминский М.П. Статистические методы планирования и обработки результатов форсированных испытаний радиодеталей: Обзоры по электронной технике. - Электронная техника, сер. 8, 1987, вып. 1.
2. Вопросы математической теории надежности. Под ред. Гнеденко Б.В. М., Радио и связь, 1983.
3. Багдонавичюс В. Статистическая проверка модели аддитивного накопления повреждений. - Теория вероятностей и ее применения, 1978, т. 23, № 2.

ПРИМЕНЕНИЕ МЕТОДОВ МНОГОМЕРНОГО ШКАЛИРОВАНИЯ В ЗАДАЧАХ ОЦЕНКИ КАЧЕСТВА ОПЕРАЦИОННОЙ РЕГУЛИРОВКИ РЕЛЕ

П.Я.Чиченков, Н.Б.Якимова

Отдельно типы малогабаритных электромагнитных реле становятся соизмеримыми с полупроводниковыми приборами, выполняющая в схемах аппаратуры более сложные функции управления и коммутации [1]. При массовом производстве реле актуальной задачей становится проблема оценки качества изделия на регулировочной операции технологического процесса. Одним из возможных подходов к задаче автоматизированной оценки качества реле в процессе регулировки является подход, основанный на использовании методов многомерного шкалирования при первичной оценке регулируемых параметров, наглядного представления данных и принятия решения на основе статистических выводов. В качестве меры близости используются прямые оценки различия. Исследуются возможности применения альтернативных оценок - матрицы совместных вероятностей и матрицы условных вероятностей, в том числе матрицы перехода. Регулировщика можно рассматривать как лицо, производящее первичную обработку данных, как лицо, принимающее решение, и лицо, выбирающее конкретную последовательность действий и перевод реле в процессе регулировки в последовательность состояний. Для визуализации данных используется программа *PAPSUN* [2].

Л и т е р а т у р а

1. Ройзен В.З. Электромагнитные малогабаритные реле. Л.: Энергоатомиздат. Ленингр. отд-ние, 1986. - 252 с.
2. Дейвисон М. Многомерное шкалирование: Методы наглядного представления данных. М.: Финансы и статистика, 1988. - 254 с.

ТЕРМИНАЛЬНЫЕ МОДЕЛИ В ЗАДАЧАХ УПРАВЛЕНИЯ КАЧЕСТВОМ ПРОДУКЦИИ

Н.Н.Швайдак, Е.Н.Румянцев, С.В.Гурин

В реализации стратегии обеспечения качества продукции значимым являются статистическое мышление способное сформировать стабильную систему управления на основе формализованного описания причинно-следственных отношений, ведущих к выпуску бракованной продукции. Принципы статистического подхода предусматривают возможности использования формализованных интерпретаций результатов контроля качества и формируют предпосылки ориентации управления на расширенные информационные представления о причинах брака, локализуемых в отклонениях и дефектах качества. Кажущееся многообразие отклонений и дефектов в совокупности с их причинами является заведомо конечным и проявляется в конкретных производственных условиях. Конечно-граничные параметры дефектов и отклонений, а также возможности их формализованного описания базируются на конкретности и счетности требований к качеству, определенных стандартами производства, что обеспечивает возможности структурного моделирования причинно-следственных отношений [1], ведущих к выпуску брака.

Традиционно, математические методы, графически задающие порядок на множествах и учитывающие характер по взаимосвязям (сетевое моделирование, древовидные структуры, сети Иисаавы) не обеспечивают полноты статистического описания причинно-следственных связей, не учитывают уровни и структуру причинного воздействия. Расширение необходимых информационных представлений о причинах дефектов и отклонений качества достигаются в рамках разработанного терминального моделирования, как инструментация, графически описывающего переходные и граничные состояния по структуре причинно-следственных связей, порождающих брак. На терминальной модели задаются следующие уровни по глубине причинно-следственных связей:

- рудиментарный уровень, как начальные элементы причины дефектов и отклонений, порождающий их проявление, сущест-

зование во времени и дальнейшее развитие;

- энтопричинный уровень, имеющий ряд переходных порядков, на которых усиливается не только проявление причины, но и растет ее вероятностное воздействие на образование брака;

- факторный уровень причинного влияния формирует непосредственный риск в проявлении брака;

- терминальный уровень задает гранично-конечное состояние в развитии причины, вызывающей выпуск бракованной продукции.

Моделирование причинно-следственных связей задает две укрупненные группы признаков, оценивающих причинное воздействие:

- конкретизирующая (время, адрес, класс) совокупность параметров для всех элементов причины, отраженных в терминальной модели, по которым формируются исходные статистические массивы;

- математически-оценочные (вероятность образования, время существования, уровень воздействия) параметры модели.

В рамках технологии терминального моделирования возможно создание автоматизированных информационных фондов, отражающих статистическую структуру причинно-следственных связей, формирующих классификационное отклонение или дефект качества по всем параметрам модели. Здесь достигается на основе применения методов статистического анализа ГЛАВНОЕ – возможность предвидеть, распознать, иметь расширенное представление о наиболее раннем проявлении элемента причины возможного брака и, устранив ее, не допустить выпуск бракованной продукции.

Л и т е р а т у р а

1. Швайдак Н.Н.. "Терминальное моделирование причинно-следственных отношений, порождающих опасности" (4.1. Структурная формализация). – Красноярск, 1988 – 48 с. (Предпринт/ВЦ СО АН СССР; № II-88).

АВТОРСКИЙ УКАЗАТЕЛЬ

- Абусев Р.А. Пермский ГУ; ст.63
- Авербух Е.А. НПО "Союзпромгеофизика", Калинин; ст.157
- Адамов С.Ю. ЦЭМИ АН СССР; ст.199,286
- Адамчик В.А. Гомельский политехи.ин.; ст.266
- Адилов Т.Х. Ин. мат. АН УССР, Ташкент; ст.11
- Алтухова Н.В. Саратовский политехи.ин.; ст.320
- Альфес М.Э. Вильнюсский ин.мат.и киб. АН Лит.ССР; ст.35
- Андрукович П.Ф. ЦЭМИ АН СССР; ст.109
- Андрюшкевич О.А. ЦЭМИ АН СССР; ст.273
- Анисимов В.В. Киевский ГУ; ст.72
- Арак Т.В. Тартуский Унив.; ст.86
- Ардаева Е.Г. Ин.гигиены труда и профзабол. АМН СССР; ст.214
- Баланина Н.Ю. Лен. ГУ; ст.148
- Балаш В.А. МЭСИ ИСЭП АПК АН СССР; ст.115,288
- Батугина Е.Н. Ин.горного дела Севера СО АН СССР; ст.268
- Баумштейн П.И. Центр.НИИ комплексной автоматизации; ст.56,137
- Белан И.Б. Ин.клин.иммунологии СО АМН СССР; ст.246
- Белоусова С.В. Киевский ГУ, СКБ "Спектер"; ст.188
- Беляев Ю.К. МГУ; ст.244
- Беляева В.А. МГУ; ст.244
- Березань Е.И. Харьковский ГУ; ст.283
- Береснева И.Б. ИСЭП АН СССР; ст.270
- Богатов К.М. Ин.гигиены труда и профзабол. АМН СССР; ст.214
- Богачев В.В. МЭСИ; ст.272
- Бондаренко В.М. Карельский филиал АН СССР; ст.102
- Бородкин Л.И. МГУ; ст.118
- Бритвихин В.А. Томский ГУ; ст.31
- Бурдаев В.П. Ин.мат.с вычисл.центром АН МССР, Кишинев; ст.190
- Буторин А.Е. Новосибирский ин. инженеров железнодорож. транс.; ст.153

Бухштабер В.М.	НПО "ВНИИ физ-тех.и радиотех.измерений", Москва; ст.111,120
Буш Х.	ВЦ НПО "Силава" Лат.ССР; ст. 201
Варварин Д.Г.	ИСЭП; ст.139
Вегнер М.К.	ЦЭМИ АН СССР; ст.199
Векслер Л.С.	НПО "ВНИИ физ-тех.и радиотех.измерений", Москва; ст.120
Великий Ю.А.	Киевский ГУ; ст.252
Верева О.В.	Ин.киб.им.Глушкова АН УССР; ст.216
Вийкманн Э.	Ин.киб.ЭССР; ст.237
Вилисмяз Ю.Ю.	Тартуский Унив.; ст.220
Виноградова В.Н.	Казанский ГУ; ст.302
Витяев Е.Е.	Ин.мат.СО АН СССР, Новосибирск; ст.192
Воскобойник В.Д.	НПО "Медбиоэкономика"; ст.250
Вылегжанин Н.Д.	ЦЭМИ АН СССР; ст.204
Высоцкая Н.В.	ст.289
Вяхи М.У.	Тартуский Унив.; ст.253
Гавсюк А.	Московский финансовый ин.; ст.194
Гамбаров Г.	Московский финансовый ин.; ст.122,194
Герасимова И.А.	ЦЭМИ АН СССР; ст.286
Гиль Л.Р.	Киргизский ГУ; ст.291
Гирко В.Л.	Киевский ГУ; ст.74,76
Гладун А.Н.	Украинский филиал НИИ Госкомстата СССР; ст.260
Горчакова Н.Ф.	Саратовский экон.ин.; ст.288
Гречко В.О.	ИК АН УССР, Киев; ст.240
Губарев В.В.	Ин.клин.иммунологии СО АМН СССР; ст.246
Губин К.	Московский финансовый ин.; ст.194
Гуревич В.А.	ст.87
Гурин С.В.	Производственное объединение "КРАСТЯЖМАШ" ст.339
Давидович М.И.	Ин.мат. АН БССР; ст.12,256
Демиденко А.П.	Ин.автоматики АН Кир.ССР; ст.322
Демиденко Е.Э.	Ин.мировой экон.и междунар.отнош.АН СССР; ст.149
Денисова И.М.	ЦЭМИ АН СССР; ст.273

Дмитриев Ю.Г.	Томский ГУ; ст.89
Добровольская Л.Г.	НПО ЦМА, Москва; ст.37
Добровольский И.Г.	НПО ЦМА, Москва; ст.37
Дробышев Ю.П.	НПО "Горсистемотехника", Киев; ст.141
Дубров А.М.	МЭСИ; ст.113
Дубровский О.В.	Харьковский ин. радиоэлектроники; ст.124
Дугина Т.Н.	МГУ; ст.244
Дучинская К.	Ин.мат. и киб. АН Лит.ССР; ст.65
Дыховичный А.А.	Киевский ГУ, СКВ "Спектр"; ст.188
Егоров С.А.	ст.202
Елейко В.И.	Львовский ГУ; ст.221
Елисеева Т.В.	Ин.клин.иммунологии СО АМН СССР; ст.246
Енюков И.С.	ЦЭМИ АН СССР; ст.204
Епишин Ю.Г.	Московский кооперативный ин.; ст.39
Ермаков С.М.	Лен.ГУ; ст.159
Ефименко С.В.	Киевский ГУ, СКБ "Спектр"; ст.188
Ефимов А.Н.	Московский ин.инженеров транспорта; ст.13
Жвиреняйте Д.	Вильнюсский ин.мат. и киб.; ст.98
Жиглявский А.А.	Лен.ГУ; ст.24
Жилинская Е.И.	МГУ ст.222
Жовтяк И.В.	ИК АН УССР, Киев; ст.240
Журбенко И.Г.	МГУ; ст.91
Закиров А.Г.	Казанский ГУ; ст.302
Заславский А.А.	НПО "Медбиоэкономика"; ст.142,250
Захарин А.М.	ИК АН УССР, Киев; ст.240
Зеленюк Е.А.	Лаб.мониторинга природной среды и климата Госкомгидромета СССР и АН СССР; ст.304
Зубенко А.А.	ВНИИ физ-техн. и радиотехн. измерений, Москва; ст.306
Зубенко Г.М.	ВНИИ физ-техн. и радиотехн. измерений, Москва; ст.306
Иванилова Т.Н.	ВЦ СО АН СССР, Красноярск; ст.126
Иващенко П.А.	Харьковский ГУ; ст.285
Ильин Е.М.	ИСЭП АН СССР; ст.300
Илюхин В.П.	Сибирский физ-техн.ин.,Новосибирск; ст.15
Йоала В.	СКБ ВТ Ин.киб. АН ЭССР, Таллинн; ст.223

- Кабанов Ю.М. ЦЭМИ АН УССР; ст.206
- Калита А.В. Ин.киб.им.Глушкова АН УССР; ст.207,216
- Каплан Е.И. Украинский геологоразведочный ин.; ст.128
- Каплинский А.И. Воронежский ГУ; ст.151
- Каролин М. Ин.киб. АН ЭССР, Таллинн; ст.209,211,218
- Карпинка Е.С. Ин.киб.им.Глушкова АН УССР; ст.216
- Кац Б.М. Украинский филиал НИИ Госкомстата СССР; ст.260
- Качинский А.В. ВНИИ гигиены и токс.пестицидов, полимеров и пласт.масс МЗ СССР, Киев; ст.308
- Кинкар Т.Ф. Тартуский Унив.; ст.213
- Киселев Н.И. ст.275
- Китаева А.В. Сибирский физ-техн.ин.им.Кузнецова; ст.41
- Кичанова Е.В. Пермский ГУ; ст.63
- Клейн А. ВЦ НПО "Силава" Лат.ССР; ст.201
- Клесов И.И. Киевский ГУ, СКБ "Спектер"; ст.188
- Клесов О.И. Киевский ГУ, СКВ "Спектер"; ст.188
- Кобзев В.Г. Харьковский ин.радиоэлектроники; ст.13,124,323
- Кожаринова Е.Е. Ин.гигиены труда и профзабол. АМН СССР; ст.214
- Колбасова А.В. МГУ; ст.277
- Колесник Т.И. Белорусский ГУ, Минск; ст.197
- Кодло Т.Х.-А. Тартуский Унив.; ст.78
- Колпаков И.П. МЭСИ; ст.255,272
- Колчинский В.И. Киевский ГУ; ст.80,188
- Королев В.Б. МЭСИ, Москва; ст.113
- Коростелев В.Г. ИСЭП АПК АН СССР; ст.143
- Котюков В.И. Новосибирский ин.инж.железнодорож.транс.; ст.153
- Кравец Т.В. Киевский ГУ, СКБ "Спектер"; ст.188
- Кравченко Г.Г. НПО "Бурение"-ВНИИКР нефть, Краснодар; ст.324
- Криволапов С.Я. МВТУ им.Баумана; ст.336
- Кузнецов В.П. Московский ин.связи; ст.155
- Кузнецов С.Е. ЦЭМИ АН СССР, Москва; ст.224,226

Кулинская Е.В.	НПО "Медбиозкономика"; ст.232
Курочкина Т.К.	Новосибирский сельскохозяйств.ин.; ст.325
Кутенков Р.П.	ИСЭП АПК АН СССР; ст.143,160
Кушнир А.Ф.	Ин.физики Земли АН СССР, Москва; ст.42
Липидус В.А.	Горьковский филиал ВНИ ИНМАШ; ст.242
Левандо Д.В.	ЦЭМИ АН СССР, Москва; ст.224
Левина Т.И.	Ин.мат. АН БССР; ст.256
Лелумезс Э.	Ин.киб.ЭССР; ст.237
Либероль Б.Д.	Харьковский ин. радиоэлектроники; ст.44
Липовецкий С.С.	ВНИПИ труда в строительстве Госстроя СССР; ст.46
Лифенко Ю.А.	Томский ГУ; ст.48
Лугачев М.И.	ст.162
Лумельский Я.П.	Пермский ГУ; ст.291
Маамяги А.	Ин.экон. АН ЭССР; ст.310,311
Мазур В.О.	Ин. гигиены труда и профзабол. АМН СССР; ст.214
Майборода Р.Е.	Киевский ГУ; ст.100,188
Майданчик А.	Московский финансовый ин.; ст.122,194
Майстренко О.В.	Киевский ГУ; ст.236
Макаров А.П.	МГУ; ст.244
Максимов Ю.Д.	Лен. политехн.ин.; ст.164
Малыгин С.	Ин.экон. АН ЭССР; ст.311
Малюга О.С.	Киевский ГУ; ст.236
Малюгин В.И.	Белорусский ГУ, Минск; ст.50
Малютов М.Б.	МГУ; ст.52
Мандель И.	Алма-Атинский ин. нар.хозяйства; ст.130
Маркова Е.В.	ВНИИНС; ст.195
Марченкова И.Б.	Белорусский НИ сан-гиг.ин.; ст.313
Марчюлайтис С.	Вильнюсский ин. мат. и киб.; ст.66
Маслак А.А.	ВНИ и технолог.ин.биол.пром.; ст.25
Маслов В.К.	НПО "ВНИИ физ-техн.и радиотехн.измерений" ст.111
Маслякова Т.В.	ИСЭП АН СССР; ст.293
Мацкевич Д.А.	Ин. мат. АН БССР, Минск; ст.256
Медведев А.В.	ВЦ СО АН СССР, Красноярск; ст.37,326

- Медвинский М.И. НПО "Горсистемотехника", Киев; ст.141
Мелс Т.Э. Ин.зоол.и бот.АН ЭССР, Тарту; ст.312
Минина Т.Р. ИСЭП АН СССР; ст.293
Минскер И.Н. Центр.НИИ компл.автомат.; ст.279
Миркин Б.Г. ЦЭМИ АН СССР; ст.132
Митоян А.А. Ереванский ин.нар.хозяйства; ст.295
Митрофанов В.Б. НПО "Медбиоэкономика"; ст.250
Морозов В.П. МЭСИ, ИСЭП АПК АН СССР; ст.255
Мучник И.Б. Ин. проблем управления АН СССР; ст.118
Мхитарян В.С. МЭСИ, ИСЭП АПК АН СССР; ст.115,328
Нагаев А.В. Ташкентский автодорожный ин.; ст.27
Наумова Е.Н. Ин.клин.иммунологии СО АМН СССР; ст.246
Никифоров А.М. Ин.гигиены труда и профзабол.АМН СССР; ст.214

Николенко Н.Г. Харьковский ин.радиоэлектроники; ст.323
Новиков И.Д. ВКНЦ АМН СССР; ст.129,156
Ольман В. Ин.киб.АН ЭССР, Таллинн; ст.211,281
Орлов А.И. Ин.гигиены труда и профзабол. АМН СССР; ст.133,324,329

Павленко Т.В. Киевский ГУ; ст.82
Павлов Ю.Л. Карельский филиал АН СССР; ст.102
Падубинскас Г. Вильнюсский ин. мат. и киб.; ст.68
Парасюк И.Н. Ин. киб. им.Глушкова АН УССР; ст.216
Перекрест В.Т. ИСЭП АН СССР; ст.145,300
Пестунов И.А. ВЦ СО АН СССР, Красноярск; ст.326
Пестунова Т.М. ВЦ СО АН СССР, Красноярск; ст.228
Петерсен И. Ин.киб.АН ЭССР; ст.237,297
Петрович М.Л. Ин.мат. АН БССР; ст.256
Пиворюнас В. Вильнюсский ин.мат. и киб.; ст.248
Пикалис В. Ин.мат. и киб. АН Лит.ССР; ст.182
Пилявская О.С. Киевский ГУ; ст.188
Пирштук И.К. Белорусский ГУ им. Ленина; ст.135
Плинер В.М. ИСЭП АН СССР; ст.29
Подколотный Н.А. ст.192
Полис Р. ВЦ НПО "Силава" Лат.ССР; ст.201
Похъченко В.И. ВНИИ гиг.и токс.пестицидов,полимеров и

	пласт.масс, Киев; ст.308
Пригарина Т.А.	ЦЭМИ АН СССР; ст.230
Провотарь А.И.	Ин.киб. им.Глушкова АН УССР ; ст.216
Пярна К.А.	Тартуский Унив.; ст.147
РAUDис Ш.	Ин.мат.и киб.АН Лит.ССР; ст.104
Раушенбах Г.В.	НПО "Медбиозэкономика"; ст.142,232,250
Репко А.А.	Киевский радиозавод; ст.323
Ростовцев П.С.	Ин.экон. и организации пром.производства СО АН СССР, Новосибирск; ст.93
Руденко О.Г.	Харьковский ин. радиоэлектроники; ст.44
Румянцев Е.Н.	Производственное объединение "КРАСТЯЖМАШ" ст.339
Руссман И.Б.	Воронежский ГУ; ст.331
Рыжаков А.П.	Томский ГУ; ст.332
Рябкова О.И.	Белорусский НИ сан-гиг.ин.; ст.313
Сапожников П.Н.	Пермский ГУ; ст.315
Сарницкая Н.П.	ВНИИ гиг.и токс.пестицидов,полимеров и пласт.масс, Киев; ст.308
Сатаров Г.А.	НПО"Медбиозэкономика", СП"Динамика";ст.232
Седова М.А.	ЦЭМИ АН СССР, ВЦ АН СССР; ст.206
Семеняк И.В.	Харьковский ГУ; ст.283,285
Сердобольский В.И.	Московский ин.электронного машиностроения; ст.234
Середа Ю.С.	НИИ механики при Горьковском ГУ; ст.17
Сильвестров Д.С.	Киевский ГУ, СКБ "Спектер"; ст.188,258
Скопецкая Н.П.	Ин.киб.им. Глушкова АН УССР; ст.216
Сливко-Кольчик С.В.	Ин.гигиены труда и профзабол. АМН СССР; ст.214
Смирнов С.В.	ст.162
Смирнова О.С.	ЦНИИ Электроника; ст.91
Снурников М.Я.	Харьковский ин. радиоэлектроники; ст.124
Соболев С.В.	Московский текстильный ин.; ст.334
Солнцев В.Н.	Лен. ГУ; ст.148
Соломоник А.С.	ст.204
Срагович Г.В.	ВЦ АН СССР; ст.54,206
Стадник О.Е.	НИИ мед.радиологии АМН СССР; ст.118

Староверов О.В.	ЦЭМИ АН СССР; ст.299
Степанов В.С.	МГУ; ст.234
Степанова М.Д.	Белорусский ГУ; ст.197
Степахно В.И.	ИМ АН УССР, Киев; ст.19,76
Сургайлис Д.	Ин.мат.и киб. АН Лит.ССР; ст.86
Тарасенко Ф.П.	Томский ГУ; ст.31
Тарасенко П.Ф.	Сибирский физ-техн.ин.; ст.95
Терещенко О.В.	ст.289
Терпугов А.Ф.	Томский ГУ; ст.332
Тескин О.И.	МВТУ им.Баумана; ст.336
Тийт Э.-М.А.	Тартуский Унив.; ст.97,259
Тийтс Т.	Ин.киб. АН ЭССР; ст.209,211,218
Титов А.С.	Ин.гигиены труда и профзабол. АМН СССР; ст.214
Товмаченко Н.Н.	Киевский ГУ; ст.236
Тоодинг Л.-М.	Тартуский Унив.; ст.262
Торговицкий И.Ш.	Центр.НИИ комплексной автоматизаций;ст.56
Траат И.К.	Тартуский Унив.; ст.84
Трошин Л.И.	МЭСИ ИСЭП АПК АН СССР; ст.115
Трусенков С.Т.	Ин.автом.и процессов упр.с ВЦ ДВО АН СССР, Владивосток; ст.317
Трусенкова О.О.	Ин.автом.и процессов упр.с ВЦ ДВО АН СССР, Владивосток; ст.317
Углова С.В.	Ин.гигиены труда и профзабол. АМН СССР; ст.214
Ульянов Г.В.	Новосибирский ГУ; ст.21
Умывакин В.М.	Воронежский ГУ; ст.151,331
Устинов Ю.К.	Томский ГУ; ст.89
Федоров В.В.	Ин.экон. и прогноза научно-техн.прогресса АН СССР, Москва; ст.58
Фример Е.В.	ИСЭП АН СССР; ст.293
Фролов С.И.	Естественно-научный ин. при Пермском ГУ; ст.59
Фролова Л.Л.	Казанский ГУ; ст.302
Хайдуков О.Ю.	Киевский ГУ, СКБ "Спектер"; ст.188
Халилеев А.А.	ЦЭМИ АН СССР; ст.199,226

- Хамдамов И.М. Ин. мат. АН УзССР, Ташкент; ст.11
- Харин Ю.С. Белорусский ГУ им. Ленина, Минск; ст.70,135
- Хачатурова Т.В. ИСЭП АН СССР; ст.300
- Хоолма М. Ин.киб.ЭССР, Таллинн; ст.237
- Цапенко М.П. Ин.клин.иммунологии СО АМН СССР; ст.246
- Цибель Н.А. ЦЭМИ АН СССР, Москва; ст.264
- Чеботарев П.Ю. Московский ин.нефти и газа им.Губкина; ст.33
- Чеботарев Н.Г. ИМЭМО АН СССР; ст.195
- Чепонис К. Вильнюсский ин.мат.и киб.; ст.238
- Черномордик О.М. НИИ гигиены труда и профзаб. АМН СССР; ст.329
- Черноусов Е.А. МЭСИ; ст.165
- Чихенков Ю.Я. Харьковский ин.радиоэлектроники; ст.338
- Чуприн С.В. Украинский НИИ металлов; ст.328
- Швайдак Н.Н. Производственное объединение "КРАСТЯЖМАШ"; ст.339
- Шевляков Г.Л. Лен. политехн.ин.; ст.116,164,319
- Шимельмиц А.Г. ЦЭМИ АН СССР; ст.199
- Шир Л.Н. ВКНЦ АМН СССР; ст.129
- Шитик А.Н. Ин.мат. АН БССР; ст.256
- Шлег Г.К. Ин.мат. АН БССР; ст.256
- Штродах О. ВЦ НПО "Силава" Лат.ССР; ст.201
- Шураков В.В. МЭСИ; ст.165
- Шурыгин А.М. МГУ; ст.23,61
- Щербина В.А. Госкомиссия по сортоиспытаниям Госагропрома СССР; ст.25
- Юровский А.Ю. Ин.гигиены труда и профзабол. АМН СССР; ст.214
- Якимова Н.Б. Харьковский ин.радиоэлектроники; ст.338

65 коп.